DOCUMENT RESUME

ED 285 954 UD 025 749

TITLE	Educational Achievement: Explanations and Implications of Recent Trends.
INSTITUTION	Congress of the U.S., Washington, D.C. Congressional Budget Office.
REPORT NO	CBO-76-0-87-1
PUB DATE	Aug 87
NOTE	118p.; For the publication "Trends in Educational Achievement," see ED 284 941.
PUB TYPE	Information Analyses (070) -- Legal/Legislative/Regulatory Materials (090)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Academic Achievement; Achievement Gains; Achievement Rating; Achievement Tests; *Data Interpretation; *Educational Trends; *Social Influences; Standards; Statistical Analysis; *Test Interpretation
IDENTIFIERS	Test Score Decline

ABSTRACT

This report analyzes possible causes of current trends in educational achievement and discusses implications for policy. It uses data described in the publication "Trends in Educational Achievement," a Congressional Budget Office study released in April 1986, which assessed data about trends in test scores. This report begins with a discussion of the current controversy about achievement and goes on to describe the methods and problems of collecting and using test scores as measures of achievement. Next the report discusses approaches to explaining achievement trends. The longest section of the report discusses possible causes of those trends, among which are the following: (1) the changing ethnic composition of the school-age population; (2) increasing family size; (3) a watering down of course content in secondary schools; (4) changes in the amount of homework done by high school students; (5) Title I/Chapter 1 programs; (6) desegregation; and (7) changes in students' use of alcohol and other drugs. Also discussed are factors which probably did not contribute to the trend, such as state graduation standards. The report then evaluates educational policies and recommends ways of improving educational achievement. An appendix is included which summarizes evidence pertaining to the contributions of specific factors to test score trends. (PS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

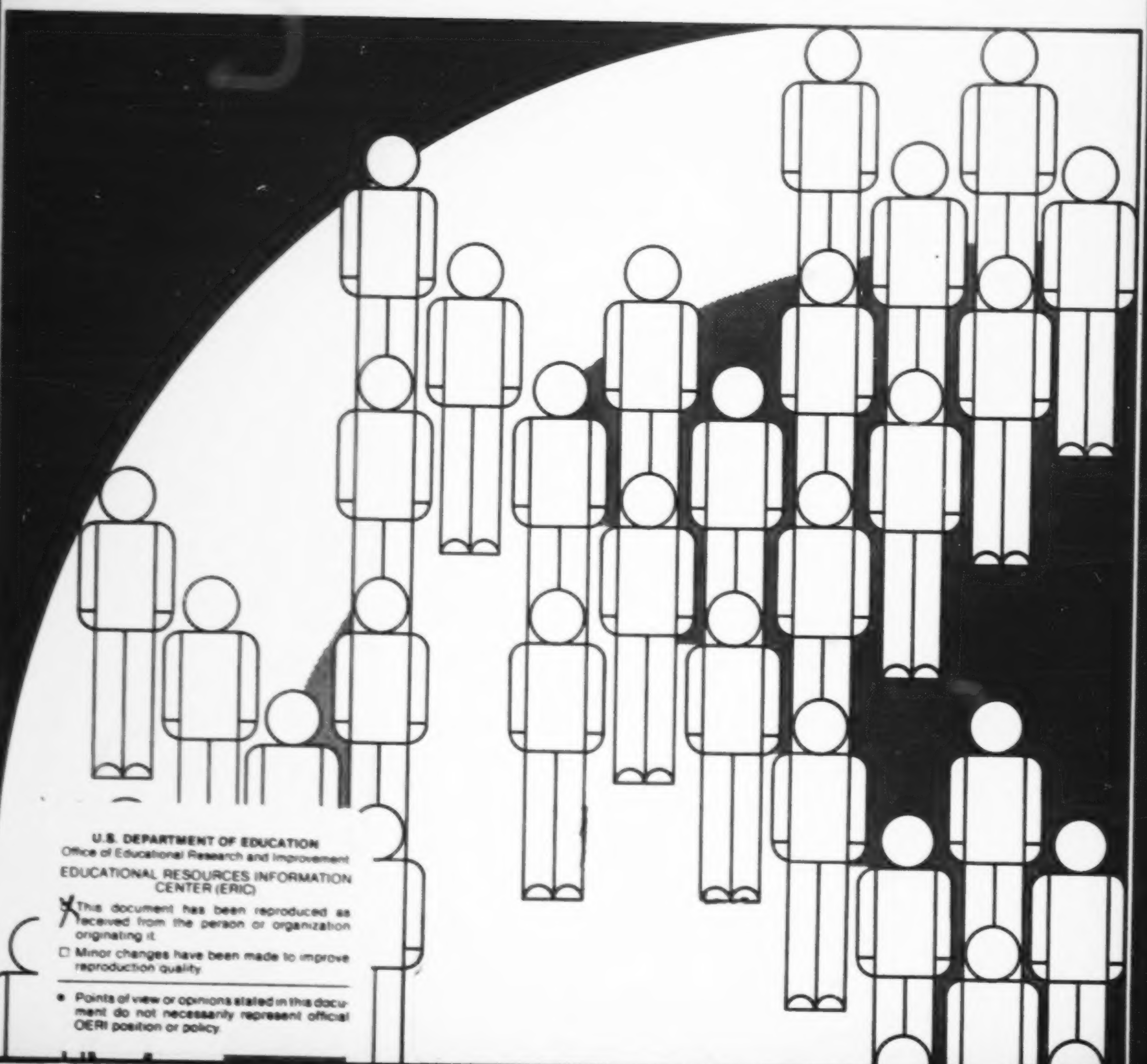
CONGRESSIONAL BUDGET OFFICE

AUGUST 1987

Educational Achievement: Explanations and Implications of Recent Trends

ED285954

BEST COPY AVAILABLE



UD025 749

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CBO STUDY

EDUCATIONAL ACHIEVEMENT: EXPLANATIONS AND IMPLICATIONS OF RECENT TRENDS

The Congress of the United States
Congressional Budget Office

76-170 0 - 87 - 1

NOTES

Except where otherwise noted, dates used in this paper are school years rather than calendar years. For example, the results of a test administered in the fall of 1979 and the spring of 1980 are both labeled 1979. As a result, the dates used here are in some instances a year earlier than those in other published sources. This discrepancy is particularly common in the case of college admissions tests and other tests administered to high school seniors, which are often labeled in other sources in terms of the calendar year in which students would graduate.

Details in the text and tables of this report may not add to totals because of rounding.

PREFACE

At the request of the Subcommittee on Education, Arts, and the Humanities of the Senate Committee on Labor and Human Resources, the Congressional Budget Office (CBO) prepared this study of trends in the educational achievement of elementary and secondary school students. The first part of the study, *Trends in Educational Achievement*, published in April 1986, presented an analysis of the achievement trends themselves. This paper assesses the causes and implications of the trends. In accordance with CBO's mandate to provide objective and impartial analysis, neither volume contains recommendations.

Daniel Koretz of CBO's Human Resources Division prepared the analysis under the direction of Nancy M. Gordon and Martin D. Levine. Sherry Snyder edited the report. Ronald Moore typed and prepared the manuscript for publication.

Many individuals and organizations contributed in various ways to this report. Among the organizations that provided extensive data, some of which is unpublished, are the Iowa Testing Programs, the National Assessment of Educational Progress, and the A. C. Nielsen Company. Special thanks are due H. D. Hoover of the Iowa Testing Programs and Lawrence Rudner of the U. S. Department of Education, who provided insightful contributions at many stages of the project. Eric H. Hanushek of the University of Rochester, and Lyle V. Jones of the University of North Carolina, offered thorough and helpful reviews.

Edward M. Gramlich
Acting Director

August 1987

CONTENTS

SUMMARY

I INTRODUCTION 1

The Context of the Current Controversy *2*
The Federal Role in Elementary and Secondary Education *3*
Recent Policy Initiatives *7*

II EDUCATIONAL ACHIEVEMENT: FACTS AND UNCERTAINTIES 9

Test Scores as a Measure of Educational Achievement *9*
Patterns in the Achievement Data *11*
Important Gaps and Inconsistencies in the Achievement Data *17*

III APPROACHES TO EXPLAINING ACHIEVEMENT TRENDS 21

Evaluating Evidence About Specific Factors *22*
Inferring Causes from General Patterns in the Test Score Data *25*

IV THE POSSIBLE CAUSES OF THE ACHIEVEMENT TRENDS 27

Three Groups of Contributing Factors *27*
The Effects of Educational, Selection, and Societal Factors *29*

V IMPLICATIONS 43

Assessing Educational Achievement *43*
Evaluating Educational Policies *45*
Improving Educational Achievement *48*

APPENDIX 53

Societal Factors *53*
  Changes in the Ethnic Composition of the Entire Cohort *53*
  Single-Parent Households *54*
  Family Size *57*
  Maternal Employment *66*
  Students' Use of Alcohol and Other Drugs *68*
  Television Viewing *69*
  Students' Attitudes and Motivation *71*
  Environmental Lead *73*
Educational Factors *76*
  Teachers' Skills and Experience *76*
  State and Local Graduation Requirements *81*
  Coursework *82*
  Minimum-Competency Testing *84*
  Textbook Difficulty *85*
  Homework *87*
  Demands for Writing *89*
  Grade Inflation *90*
  Educational Programs for Disadvantaged Students *91*
  Desegregation *95*
Selection Factors *98*
  Retention *99*
  Self-Selection *102*

TABLES

1. The Effects of Various Educational, Societal, and Selection-Related Factors on Recent Trends in Test Scores 32

A-1. Percent of High School Seniors Reporting Various Amounts of Time Spent on Homework Per Week, 1971 and 1979 School Years 89

A-2. Trends in SAT-Verbal Scores and Reading Comprehension Among All High School Seniors, and SAT Candidates, 1959 to 1971 104

FIGURES

1. Shares of Elementary/Secondary Education Funding by Level of Government 5

2. Iowa Average Test Scores, Grades 5, 8, and 12, Differences from Post-1964 Low Point 14

A-1. Average Total SAT and Average Birth Order (By year of birth) 59

A-2. Television Viewing by Children, by Age (Average hours per week) 71

A-3. Percent of 16- and 17-Year-Olds Enrolled Below College Level (All ethnic groups combined) 101

A-4. Percent of Age Group in Modal Grade (Three-year moving averages) 101

BOX Have High School Test Scores Stopped Rising? 19

SUMMARY

The educational achievement of American elementary and secondary school students has been the focus of unusually intense scrutiny for several years. Strong public concern has been accompanied by extensive and continuing efforts at all levels of government to improve the public educational system.

Scores on standardized achievement tests have played a central role in this debate. Few issues were as critical to kindling the debate as was a growing public awareness that the test scores of American students declined markedly during the 1960s and 1970s and compared poorly with those of students in other countries. Many of the recent educational policy initiatives, such as stiffer standards for graduation from high school, were intended to counter these trends or to offset some of the factors (lax academic standards, in this case) that were presumed to have caused them. Moreover, many initiatives have increased the use of testing--of teachers as well as students--not only to measure achievement, but also to improve it.

Given the importance currently afforded scores on standardized tests, a careful appraisal of trends in test scores and their causes has significant implications for educational policy. *Trends in Educational Achievement,* a Congressional Budget Office study released in April 1986, assessed currently available data about trends in test scores and described some of the important limitations of standardized tests. This report analyzes possible causes of those trends and discusses implications for policy.

CURRENT INFORMATION ABOUT EDUCATIONAL ACHIEVEMENT

The existence of a sizable drop in test scores during the 1960s and 1970s has been well known for some time. The decline was remarkably pervasive, affecting many different types of students in most grades, in all regions of the United States, in Catholic as well as public schools, and even in Canadian schools. The drop was apparent in the results of different kinds of tests covering many subject areas. The deterioration was greater among

older students than in the early grades and affected higher-order skills such as reasoning and problem-solving more severely than more basic, rote skills.

The decline in scores was followed immediately by a widespread and significant rise. Perhaps because of the prominence of tests administered to senior-high students (for example, the Scholastic Aptitude Test, or SAT), many observers have mistakenly believed that the upturn did not start until the beginning of this decade (when SAT scores began to increase) and that it has been relatively inconsequential. Examination of a broader range of test data, however, shows that the upturn actually began by the mid-1970s and has been sizable. On certain tests administered to young children, for example, the upturn has more than overcome the previous decline.

Underlying the confusion about the timing of the upturn is a "cohort pattern" in the test scores that is central to understanding the possible causes of these trends. A cohort pattern is a change that affects children born in the same year, rather than children of various ages in school together in a given year (known as a "period effect"). The upturn typically began within a few years of the cohorts of children who were born in 1962 or 1963 and entered school in the late 1960s. The rise in scores first became apparent in the mid-1970s, when those children were in the middle elementary grades, and gradually moved into the higher grades as they progressed through school. Since then, successive cohorts of students have typically scored progressively higher. The lesser size and later onset of the rise in scores in the higher grades appears largely to reflect the smaller number of improving cohorts to have reached that level.

Several other variations in the trends are noteworthy. Black students and probably Hispanics have gained appreciably relative to their nonminority peers, although the gaps in scores between minority and nonminority groups remain large. The data also suggest that relative gains were made by students in schools with high minority enrollments and in disadvantaged urban communities.

Even though the recent rise in test scores has been substantial, the average level of performance on some tests remains well below what many educators would consider acceptable. Serious deficiencies can be found in all levels of skills, from the most rudimentary to the advanced. Moreover, many of these weaknesses will undoubtedly hinder students in their life outside of school. A disturbingly large proportion of American students, for example, are still unable to apply fundamental skills, such as simple mathematics, to situations encountered in everyday life.

GAPS IN CURRENT INFORMATION ABOUT EDUCATIONAL ACHIEVEMENT

Although considering a wide array of test data adds to the information that can be provided by one or a few tests, it also reveals a number of unanswered questions.

While some uncertainties simply reflect a scarcity of relevant data, others have arisen because existing tests--including those of high quality--sometimes provide inconsistent answers to even basic questions about educational achievement. For example, tests offer widely divergent estimates of the relative severity of the trends in different subject areas. Similarly, there are two recent nationally representative assessments of regional differences in achievement trends: one found particularly favorable trends in the South, while the other indicated a decline in the South that was comparable to or worse than that in other regions.

Another, potentially very important discrepancy among tests concerns the performance of the cohorts that have entered school in the last few years. While there is little reason to doubt that cohorts that have recently produced gains in the lower grades will continue to raise average scores as they progress through school, it is not clear whether incoming cohorts are continuing to outperform those that preceded them. Some tests show continuing gains in the lowest grades, while others suggest stagnation. Resolution of this question, which is important to any evaluation of the current wave of educational policy initiatives, will require information from additional tests administered over the next several years.

Such inconsistencies point to a critical, but widely ignored, limitation of standardized tests: even the best of current tests are only incomplete proxies for educational achievement. Most tests measure only some of the many skills required to master a broad subject area such as mathematics, for example. Consequently, the results of tests can differ from each other, often in ways that are unanticipated and difficult to explain. Moreover, important skills such as the ability to write well are difficult to assess using current standardized tests, and even data from several tests can yield inadequate information about them.

CAUSES OF THE ACHIEVEMENT TRENDS

Although a large number of diverse factors have been suggested as causes of the recent trends, many analysts are confident that one or a few factors can account for much of the change shown by test scores over the past two decades. Moreover, many analysts believe that factors of a single type are responsible for those changes. The majority of them looks among educational factors for an explanation, while a smaller and less influential group expects the answer to be found in noneducational factors such as demographic trends and changes in students' use of alcohol and other drugs.

The available evidence, however, paints a much more complicated picture. The trends most likely resulted from the combined effects of numerous factors, both educational and noneducational. Moreover, to the extent that estimates are feasible, the individual contributions of those factors were typically modest. Two factors whose effects can be relatively well estimated and that appear to have made particularly substantial contributions to the trends--the changing ethnic composition of the school-age population and increasing family size--could each account for at most a fifth to a fourth of the total change in scores during portions of the achievement decline. The contributions of some other factors, while more difficult to estimate, appear to have been considerably smaller. Even taken together, the factors examined in this study provide only a partial explanation of the trends, and the limitations of the available data make it likely that any explanation will remain incomplete.

Perhaps because of the extensive attention paid to high school tests, many analysts who expect the achievement trends to have educational causes look to the late 1960s and 1970s--when the test scores of senior-high students were falling--for policies that might have caused the decline in scores. Similarly, many expect that the causes of the subsequent upturn can be found in the policies of the 1980s and perhaps the late 1970s.

While there is some truth in this view, it too is simpler than the data warrant. Some of the educational changes that contributed to the achievement trends were probably consistent in timing with trends in scores in the lower grades, not with scores at the senior-high level. The cohorts that produced the upturn in test scores entered school beginning in the late 1960s, and their improved performance was evident during their elementary school years. Thus, educational practices as early as the late 1960s and early 1970s--at least in elementary schools--might also have contributed to the rise in scores.

The factors that remain as plausible causes when systematic evidence is examined include a number of educational factors that often arise in the debate about achievement trends. A watering down of course content in secondary schools might have contributed to the decline in scores and might help account for the greater severity of the decline in the higher grades. Changes in the amount of homework done by high school students, though relatively modest, might have contributed to both the decline and the subsequent upturn. Chapter 1 (the federally funded compensatory education program) could have contributed modestly to the relative gains of black and Hispanic students. Desegregation also might have contributed to the gains of blacks but apparently not to those of Hispanics, since the schools that Hispanics attend have become more segregated, not less.

The noneducational factors that could have contributed to the trends include some that are widely discussed and others that have received little notice in this context. Changes in family size that accompanied the baby boom and baby bust, which have received extensive attention, probably contributed moderately to both the decline and the upturn. Changes in the ethnic composition of the student body could account for perhaps a tenth to a fifth of the decline in test scores during the 1970s but probably impeded the rise in scores somewhat. Changes in students' use of alcohol and other drugs might have contributed to both the decline and the upturn and, like changes in coursework, might help explain the greater decline in the higher grades. A decrease in exposure to environmental lead--often discussed as an influence on children's health and cognitive functioning but rarely noted as a possible cause of trends in test scores--might have contributed in small measure to the upturn.

The list of factors that probably did not contribute significantly to the trends is more surprising, because it too includes factors that have gained widespread credence as possible causes. State graduation standards, for example, did not change significantly between 1974 and 1979 and therefore appear not to have contributed directly to the latter half of the achievement decline, and systematic data about requirements in earlier years are not available. Several commonly cited noneducational factors also do not weather close scrutiny. Whatever their effects on achievement in general, for example, neither television viewing nor the growing proportion of students living in single-parent households appear to have caused any significant share of the decline in test scores; the former did not change in ways that would have contributed to the trends in test scores, and the latter changed too little to have mattered in this context.

Finally, a number of commonly cited factors cannot be evaluated because existing data are inadequate. This gap in information is serious, because some of the factors that cannot be assessed have been important in the current debate and might have a substantial influence on test scores. These factors include local graduation requirements and students' motivation and attitudes toward education.

IMPLICATIONS

The analyses reported here have broad implications for assessing the condition of educational achievement and for formulating and evaluating educational policies.

Gauging the Condition of Educational Achievement

Because the currently available data leave important questions unanswered, additional national data from educational tests would clearly be helpful in assessing the achievement of American students.

The analysis in this report, however, argues strongly against relying solely on a single "national achievement test" for this additional information. A more reliable and informative, though costlier, alternative would be to maintain a number of tests, which ideally would vary in content and format. A comparison of several tests is often necessary to discern which results are consistent enough to provide a sound basis for policy, as evidenced by the several important instances in which the National Assessment of Educational Progress has yielded conclusions that are inconsistent with other data, and the wide variation in the results shown by other tests. Moreover, disparities in the results of different tests can themselves provide significant information. Because tests often stress different types of knowledge and skills, divergence in their results can reveal important facts about students' mastery of various aspects of a subject area.

For certain purposes, it would be critical to collect information about pertinent educational and noneducational factors, such as demographic trends and dropout rates, to accompany data from additional educational tests. Though costly to collect, such information would be important because the extent to which trends in test scores should be seen as real changes in students' achievement depends on the mix of factors responsible for them. At one extreme, trends in test scores attributable to educational factors, such as improved curricula, represent true changes in achievement. At the other extreme, trends in test scores that result from selection factors--that is, from changes in the selection of students to be tested--usually cannot be seen as actual changes in achievement. A drop in average test scores attributable to a decline in the dropout rate, for example, or to an increase in the number of less able students taking an optional college admissions test signifies nothing about the level of educational achievement of the school-age population as a whole. In between these two extremes are trends caused by societal factors--that is, noneducational factors other than selection changes. Such trends often would be seen as real changes in achievement, but their interpretation can vary depending on the factors involved and the question being addressed.

Evaluating Educational Policies

Trends in average test scores have become a common criterion for gauging the effectiveness of educational programs. The link between trends in test scores and educational policies, however, is far less straightforward than many people assume. Even when test data are sufficient to provide reliable information about students' achievement, they can lead to erroneous inferences about the effectiveness of educational programs.

Simple trends in test scores--that is, whether test scores are rising or falling--in themselves do not indicate whether policies are effective. Because many factors of different types (educational, societal, and selection-related) influence test scores, effective policies can be accompanied by falling scores, and rising scores can accompany policies that are actually detrimental. Accurate evaluation of a policy requires information on how trends have been deflected from the course they would have followed in the absence of that policy.

In the next few years, for example, simple trends in test scores will in many instances overestimate the effectiveness of educational policy initiatives because the current rise in scores antedates many of these policies and might well have continued in their absence, at least in the higher grades. In addition, the current emphasis on testing is likely to increase the extent to which teachers "teach to the test"--that is, tailor instruction specifically to raise scores. Regardless of whether increased teaching to the test is desirable, it is likely to make trends in test scores a distorted proxy for achievement.

In certain circumstances, however, simple trends in test scores will underestimate the effectiveness of educational initiatives. For example, scores may be depressed in districts undergoing unusually rapid demographic changes even if the policies carried out during that time are beneficial. Similarly, successful efforts to lower the dropout rate are likely to depress average scores.

Improving Educational Achievement

Many people have used trends in test scores and assumptions about their causes not only to formulate new educational policies, but also as a basis for presuming their effectiveness. Some assume that a few key factors that caused the decline of the 1960s and 1970s can be identified and that reversing those factors will cause scores to rise as markedly and as pervasively as they fell during those years.

Far from identifying a few key factors, however, this study suggests that changes in many, diverse educational factors might well be necessary to bring about increases in achievement as pervasive and large as the decline of the 1960s and 1970s. The individual contributions of educational factors to the recent trends were apparently modest. Moreover, since non-educational factors caused a sizable share of the change, even the effect of all educational causes combined, including factors not assessed here, fell substantially short of the total change in scores observed during those years. Thus, to bring about an increase as large and widespread as the decline would require a more powerful mix of educational changes than that which contributed to the decline.

This study thus suggests searching broadly for educational factors that might improve achievement. Focusing on factors that contributed to the trends of the recent past--for example, changes in the amount of homework assigned--might be productive. But the effects of those factors may be more modest than hoped, and limiting the search to them could exclude other factors of equal or greater importance. Factors whose contributions to the recent trends cannot be appraised for want of data, for example, include some--such as students' attitudes, demands for writing, and local graduation requirements--that might exert a powerful influence on students' learning. Even certain factors that apparently did not contribute to the recent trends--specifically, those, such as state graduation requirements, that did not change sufficiently during the relevant years--might also be important in the future.

Indeed, the results of this analysis suggest that the effectiveness of the current wave of initiatives should not be presumed on the basis of assumptions about what caused past trends. In many ways, the initiatives are more appropriately seen as an experiment than as a clear-cut response to the trends of the past two decades, and careful evaluation will be needed to assess their effects--both positive and negative.

Even though this study did not uncover the small number of key factors that many people would like to find, it does have several implications for the design of future initiatives. First, initiatives aimed primarily or entirely at the secondary level--for example, stiffened graduation requirements--even if beneficial, will miss an important part of the problem. The trends evident in the higher grades were also apparent in lower grades, and many of the skills in which deficiencies are particularly striking are taught in elementary and junior high schools.

Second, the data highlight the importance of improving higher-order skills, such as reasoning and problem-solving, at all grade levels. Even though many rudimentary skills must be strengthened, policies that focus too much on rote skills and too little on reasoning and problem-solving will fail to address, and might even worsen, problems with higher-order skills that the test score data reveal to be particularly severe.

Finally, this analysis also suggests the need to focus on the performance of certain traditionally low-scoring groups but reaches no conclusions about the form that such initiatives should take. Although certain of these groups--for example, black students--have made appreciable gains, their level of achievement is still far below the national norm. The factors commonly advanced to account for these relative gains--desegregation and federally funded compensatory education--probably account for some of the improvement but leave much of it unexplained. Given the lack of an explanation for the rest of the improvement, there is a real danger that policies that were beneficial in this respect could be inadvertently discarded or undermined in the process of altering educational policy more generally. Only careful monitoring of the effects of the current wave of initiatives on the education of these students will clarify which of the changes further their recent gains and which erode them.

CHAPTER I

INTRODUCTION

Concern about the quality and effectiveness of American elementary and secondary education has been unusually intense for several years, perhaps greater than at any time since the Sputnik-inspired reform era of three decades ago. This concern has had many expressions: extensive coverage in the press, numerous influential reports on the status of education, and widespread political attention and efforts at all levels of government to improve the educational system.

Measures of educational achievement, particularly scores on various types of standardized tests, have played a key role in this ferment. One of the wellsprings of the debate was a growing public awareness that by many measures, the educational achievement of American students dropped considerably during the 1960s and 1970s, and that it compares unfavorably with the performance of students in some other countries. This information from educational tests, and the abundant hypotheses about the causes of these deficiencies in performance, have played a central role in forming the current spate of educational initiatives at all levels of government. Many of these initiatives are responses to problems revealed by such tests, and test scores have been cited as being a part of their rationale.

The influence of tests on both educational practice and public discussion has also increased. Many of the recent educational initiatives entail using these tests more and giving them greater importance. Examples include increased reliance on tests as prerequisites for high school graduation and the use of tests to screen potential teachers. In addition, Americans appear to have come increasingly to judge the quality of their schools by the results of achievement tests--a trend that is apparent from the local level to the national. Indeed, standardized tests have become a sort of national report card. Local newspapers routinely publish comparisons of schools in terms of the average test scores of their students. At the national level, the Department of Education has begun publishing periodic comparisons of the educational systems of the 50 states, highlighting the average scores on college admissions tests of the students in each state who take those tests.

Over the past year or so, more positive trends in educational achievement have gained increasing attention. It is now widely known that the decline of test scores during the 1960s and 1970s has ended and has been followed by a substantial rise. Although the more favorable recent trends in test scores have not yet affected the current wave of educational policy initiatives in a way comparable to that of the preceding decline, they too have been incorporated into the national report card and have been cited by many observers as an indication that the educational system is improving.

The current importance attached to test data makes it critical to appraise recent trends in test scores accurately and to evaluate explanations of those trends carefully. *Trends in Educational Achievement*, a Congressional Budget Office study released in April 1986, assessed much of the available information about trends in test scores and described some of the important characteristics and limitations of common tests. (Several conclusions of the earlier report that are crucial to an understanding of this paper are summarized here in Chapter II.) This report supplements the earlier one by analyzing possible causes of trends in test scores. Some of the most common or influential explanations are evaluated by assessing their consistency with the broad array of test data analyzed in the earlier report and with other, independent evidence. In addition, this report explores the implications for policy of both the trends and their causes.

THE CONTEXT OF THE CURRENT CONTROVERSY

While elementary and secondary education remains primarily a state and local responsibility, it is a truly national concern. Debate about education policy frequently stresses questions of national interest, such as the impact of education on the productivity of the nation's work force and consequently on the international competitiveness of the American economy and the nation's security. The current debate has been shaped by the reports of numerous national commissions, the National Governors' Association, the Council of Chief State School Officers, and the Department of Education, as well as other regional and national groups. Moreover, many of the recent changes in educational policy and practice have been national in scope, as many states followed common paths in making independent decisions.

The themes of the current controversy--and participation in the debate by members of the Congress and the Administration--have longstanding historical precedents. This continuity is perhaps clearest in the concern about the possible consequences of education for the productivity of the work force and the competitiveness of the American economy, which has been a recurring theme in legislation and in debates about educational policy at least since the turn of the century. For example, one of the aims of the Smith-Hughes Act of 1917, which established federal support for vocational education, was to improve the skills and productivity of the work force in response to international competition. 1/ That act, which is commonly acknowledged as the first federal program of categorical aid to elementary and secondary education, is still funded today.

More recently, the report of the National Commission on Excellence in Education, *A Nation at Risk*, stated that "Our once unchallenged preeminence in commerce, science, and technological innovation is being overtaken by competitors throughout the world. This report is concerned with only one of the many causes and dimensions of the problem, but it is the one that undergirds American prosperity, security, and civility." 2/ A particularly influential report, *A Nation Prepared: Teachers for the 21st Century*, issued by the Carnegie Forum on Education and the Economy, asserted that "America's ability to compete in world markets is eroding....As in past economic and social crises, Americans turn to education. They rightly demand an improved supply of young people with the knowledge...and skills to make the nation once again fully competitive." 3/

Concern has also been voiced about the perceived failure of the educational system to challenge the nation's most able students. This too has been a recurrent theme and can be traced back at least as far as the 1893 report of the "Committee of Ten," considered by some historians to be the first major national report on the high school. This concern has been the focus of several recent congressional initiatives.

THE FEDERAL ROLE IN ELEMENTARY AND SECONDARY EDUCATION

The federal government has always played a more limited role in elementary and secondary education than have states and localities. Together, states

---

1. Carl F. Kaestle and Marshall S. Smith, "The Federal Role in Elementary and Secondary Education, 1940-1980," *Harvard Educational Review*, vol. 54, no. 4 (November 1982), pp. 384-408.

2. National Commission on Excellence in Education, *A Nation at Risk* (Washington, D.C.: Government Printing Office, 1983), p. 5.

3. Task Force on Teaching as a Profession, *A Nation Prepared: Teachers for the 21st Century* (Washington, D.C.: Carnegie Forum on Education and the Economy, May 1986), p. 2.

and localities provide most of the funds for public education--over 90 percent, by the most common accounting--and they retain control over most aspects of educational policy and practice. 4/ Decisions about teacher certification, curricula and course requirements, and achievement testing, for example, all rest with state and local governments.

Nonetheless, the roles of the Congress and the Administration have at times been more significant than the relatively small federal share of funding might suggest. In certain areas, such as the education of handicapped or educationally disadvantaged students, the federal role is central. The federal government also influences elementary and secondary education by means other than the funding of educational services; it assumes major responsibility for collecting and disseminating educational information and statistics.

Changes in the Scope of Federal Aid to Education

During the decades following World War II, federal aid to education grew markedly. Until the mid-1940s, federal contributions had accounted for less than 1.5 percent of total revenues for public elementary and secondary education. The federal share then rose markedly for about three decades, reaching a peak of almost 10 percent in the late 1970s (see Figure 1). Since then, the federal share has fallen considerably. In the 1984-1985 school year, federal contributions of nearly $9 billion constituted about 6.5 percent of total revenues for education--the smallest share in two decades.

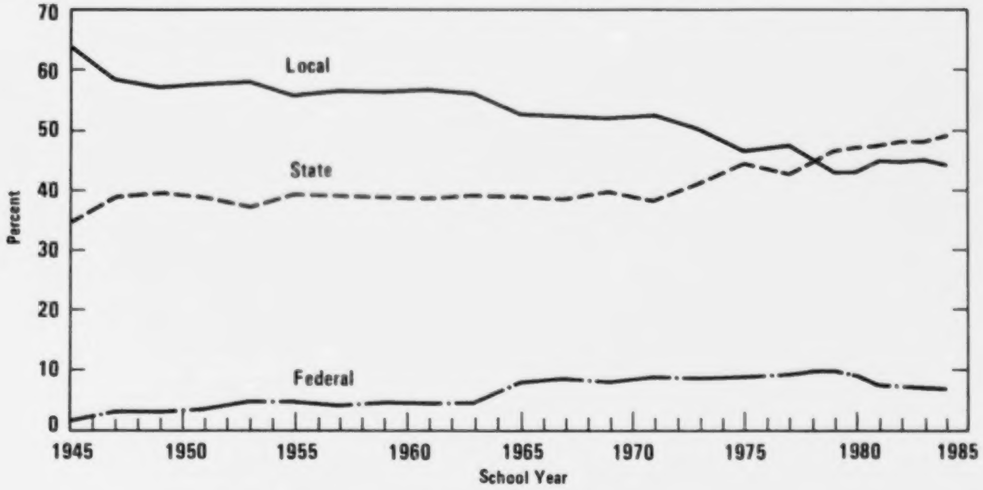
The postwar increase in the federal share of education revenues reflected major qualitative changes in the goals of federal involvement. Until the 1950s, federal aid for education was devoted to only a few purposes, such as vocational education, the education of Native American children, and fiscal assistance to localities affected by federal installations. Moreover, in 1950, more than half of all federal aid was provided for the school lunch program, not for specifically educational programs.

Since 1950, a variety of laws have broadened the scope of federal assistance for education. The National Defense Education Act of 1958 (NDEA), for example, authorized various activities intended to improve instruction in mathematics, sciences, and foreign languages. The Elementary and Secondary Education Act of 1965 (ESEA, Public Law 89-10), which produced the large increase in federal funding in the mid-1960s, authorized

---

4. Department of Education, Office of Educational Research and Improvement, *Digest of Education Statistics, 1985-86* (February 1986), Table 69.

Figure 1.
Shares of Elementary/Secondary Education Funding by Level of Government



SOURCE: Office of Educational Research and Improvement, *Digest of Education Statistics, 1987* (Washington, D.C.: Department of Education, 1987).

a wide range of programs, including the program of compensatory education that--as Chapter 1 of the Education Consolidation and Improvement Act of 1981--remains the largest single source of federal funds for elementary and secondary education.

Although these programs represented substantive changes in the character of federal aid, many of the rationales behind them echoed earlier concerns. The statement of the purpose of the NDEA, for example, noted that:

> The Congress hereby finds and declares that the security of the Nation requires the fullest development of the mental resources and technical skills of its young men and women. The present emergency demands that additional and more adequate educational opportunities be made available. . . . 5/

---

5. Public Law 85-864, Section 101 (72 Stat. 1580). See also Kaestle and Smith, "The Federal Role in Elementary and Secondary Education," p. 393.

Similarly, although the main purpose of the ESEA was to improve the opportunities open to disadvantaged students, it too reflected the concerns of Smith-Hughes and the NDEA--the effect of inadequate education on the nation's well-being. 6/

Federal Support of Educational Statistics and Research

In addition to providing financial support for certain educational services, the federal government has long been involved in elementary and secondary education by generating, collecting, and disseminating statistics and research about education. The U.S. Department of Education was established in 1867 primarily to gather educational statistics, and that function has continued without interruption to the present. The Bureau of the Census also collects statistical information about students and school districts.

This role has grown substantially in recent years. The Education Amendments of 1972 (Public Law 92-318), for example, established the National Institute of Education, now a part of the Office of Educational Research and Improvement, which has been a major source of funding for research on education. Federal efforts to gather or disseminate educational information have also accompanied programs of direct financial support of educational services. Several current proposals would further expand the federal role in gathering educational information. The report of the Secretary of Education's panel on improving the assessment of student performance, for example, recommended greatly expanding the National Assessment of Educational Progress to permit state-by-state comparisons of student achievement. 7/

Although information-related activities absorb only a modest share of federal funding for elementary and secondary education, the federal funds provide a large part of the resources for carrying them out. 8/ In a number of cases, the data generated by the federal government are unique. For

---

6. *Elementary and Secondary Education Act of 1965*, Report No. 89-143, House Committee on Education and Labor, to accompany H. R. 2362, 89:1 (1965), pp. 1448-1449.

7. Lamar Alexander, H. Thomas James, and others, *The Nation's Report Card: Improving the Assessment of Student Achievement* (Washington, D.C.: Office of Educational Research and Improvement, 1987).

8. For example, in fiscal year 1986, funding for the Office of Educational Research and Improvement, which accounts for a large share of federal support for educational statistics and research, totaled about $64 million--about three-tenths of one percent of the Education Department's appropriation of $19.5 billion.

example, all of the nationally representative data on educational achievement test scores used in this and the preceding report were federally funded. Moreover, the impact of those data in many cases is far greater than their relatively small share of funding might suggest, because they can influence educational policy and practice at all levels of government.

RECENT POLICY INITIATIVES

The intensity of the current debate about educational achievement has been matched by the abundance of policy initiatives proposed--and, in many cases, already carried out--at all levels of government. Many states and localities have instituted sweeping policy changes affecting a wide range of educational practices. Common initiatives have included increased coursework requirements for high school graduation, expanded programs of student testing, changes in standards for teacher certification, and modifications of rules for teacher compensation.

The federal responses have also been diverse, but many have been consistent with past federal efforts. The Administration has emphasized its role of disseminating information in its attempts to alter education policy and practice--for example, by issuing comparisons of the states' educational policies and outcomes. Some of the legislation considered by the Congress has followed traditions established by the NDEA, the ESEA, and Smith-Hughes. The Education for Economic Security Act of 1984 (Public Law 98-377), for example, followed the path of the NDEA in attempting to strengthen instruction in mathematics and science. Provisions with similar goals are also included in the trade bills passed by both Houses during the first session of the 100th Congress and currently awaiting conference--H.R. 3, the Trade and International Economic Policy Reform Act of 1987, and S. 1420, the Omnibus Trade and Competitiveness Act of 1987. Following in the tradition of the ESEA were the Job Training Partnership Act Amendments of 1986 (Public Law 99-496), which required that remedial education be included in certain federally funded training programs, and S. 1420, which would provide funds for a secondary school basic skills program and a dropout prevention program. In the tradition established by Smith-Hughes, H.R. 3 would also provide additional support for vocational education.

Trends in educational achievement and their presumed causes have served as rationales for many of these initiatives. Some initiatives--for example, efforts to strengthen mathematics education--focus on areas in which students' performance has shown particularly serious weaknesses or especially severe deterioration. Other initiatives, such as increases in graduation requirements, are intended to alter aspects of policy and practice that have been suggested as causes of the decline of the 1960s and 1970s, or to augment policies that might have contributed to the subsequent rise in scores.

Recent achievement trends represent only one basis for educational policy changes. Changing a particular practice might prove beneficial, for example, even if that practice--contrary to common views--did not contribute appreciably to the decline of test scores. For instance, the much discussed decline in the SAT scores between 1972 and 1979 of individuals expecting to become teachers occurred too late to have contributed appreciably to the decline in students' test scores, but that fact says nothing about the influence of teachers' academic skills on students' achievement more generally. Nonetheless, as long as the trends and their presumed causes are put forward as a justification of policy changes, it is important to evaluate the consistency between policies and these trends. Assuming greater consistency than actually exists can misdirect policy in numerous ways. It can lead to unwarranted presumptions about the effectiveness of policy initiatives, and it can obscure the importance of other factors that are less commonly viewed as being linked to the trends of the recent past.

CHAPTER II

EDUCATIONAL ACHIEVEMENT: FACTS AND UNCERTAINTIES

Because many current educational initiatives are responses to recent trends in educational achievement or to their possible causes, it is crucial to understand what the available data indicate about the achievement of elementary and secondary school students. This chapter summarizes some of the most important patterns that emerge when a wide array of data about educational achievement is examined. It is largely adapted from *Trends in Educational Achievement,* which provides more detailed information and more fully explains the limitations of existing test data.

TEST SCORES AS A MEASURE OF EDUCATIONAL ACHIEVEMENT

Current data on educational achievement are more complex, varied, and ambiguous than many observers realize. That complexity alone signals a need for caution in reaching conclusions about the condition of education, in considering possible explanations of recent trends, and in drawing inferences about appropriate policy responses.

The current debate about educational achievement was sparked by and focuses primarily on the results of standardized tests, such as college admissions tests, minimum-competency tests, and "norm-referenced" tests (tests that rate students by comparing their performance to that of other students, rather than to an absolute criterion of achievement). The debate, in turn, has prompted the burgeoning use of tests and a reliance on their results as indicators of the condition of education. Given this pivotal role of standardized tests, the strengths and limitations of test scores as an indicator of achievement are critically important.

The advantages of certain tests are considerable and apparent. The scoring of standardized tests can be free of much of the subjectivity that plagues alternative measures, such as teachers' grades. If designed and scored appropriately, tests can provide information about changes in achievement over time. Tests can also be tailored to address a wide variety of specific questions, such as the extent of progress among certain groups of students or in subject areas of particular importance.

The limitations of test scores, while less apparent, are also considerable and must be recognized. Perhaps most important, test scores are not synonymous with educational achievement; rather, a given test is usually only an incomplete proxy for the comprehensive measure of achievement that one would ideally want. Most tests can tap only a subset of the many, highly disparate skills subsumed by a subject area such as mathematics or American history. When the skills being tested are specific and narrowly defined--for example, facility with algorithms for subtraction--a test can be a reasonably close proxy. The concerns of educational policy are rarely that narrow, however. Policy debate is more likely to focus on mathematics, for example, than on subtraction. Assessing these broader areas of achievement forces important trade-offs in the design of tests. 1/

In addition, some of the skills and attitudes that schools strive to foster are difficult to gauge using standardized tests, and the assessment of students' performance can be distorted by the scarcity of information about these characteristics in the available test data. For example, the ability to write cogently is hard to assess because evaluating writing samples is both laborious and subjective, particularly in comparison with multiple-choice tests. As a result, large-scale, direct assessments of writing ability (as opposed to multiple-choice tests of language usage and writing mechanics) have been relatively uncommon until recently and have had comparatively little influence on public perception of achievement trends. Other attributes that schooling attempts to develop may be even more difficult to assess, such as an interest in reading, mastery of certain types of reasoning, and the ability and propensity to apply skills developed in school to very different and perhaps unstructured problems encountered out of school.

Another limitation of test scores as an indicator of achievement is that even similar tests can yield markedly different results. Indeed, one of the most serious mistakes made by some analysts attempting to explain recent achievement trends--or to draw implications for policy--has been to assume that patterns evident in the scores of one test will appear in

---

1. Moreover, the range of subject matter need not be very broad to force important trade-offs. One recent study of fourth-grade mathematics, for example--a subject with relatively little curricular variation--found sizable differences in the content of commonly used tests. See Donald J. Freeman, Theresa M. Kuhs, Andrew C. Porter, Robert E. Floden, William H. Schmidt, and John R. Schwille, "Do Textbooks and Tests Define a National Curriculum in Elementary School Mathematics?" *The Elementary School Journal*, vol. 83, no. 5 (May 1983), pp. 501-513.

others as well. Some of the patterns that have been prominent in the recent debate about educational policy do not appear consistently when a wide array of tests are considered.

Given that tests are incomplete proxies for comprehensive measures of achievement, some discrepancies in their results should be expected, and some of the factors that contribute to the variation in results are known. Choices made in designing the tests, for example--decisions about content, emphasis, and test format--can cause the results of tests to vary. Results can also differ because of seemingly arcane technical details. For example, the answer to the key question of whether trends in achievement have been more favorable among low-achieving students than among their high-achieving peers varies depending on how the test scores are scaled and reported. Still, some important discrepancies in the results of major tests remain unexplained.

PATTERNS IN THE ACHIEVEMENT DATA

The available data from standardized tests paint a mixed picture of the achievement of elementary and secondary school students: some aspects of the data are encouraging, while others are profoundly disturbing. This duality is especially evident when one considers both the levels of achievement shown on various tests and the trends in achievement over time. For example, promising trends can appear even when average scores remain distressingly low.

The Decline in Test Scores

The sizable drop in test scores during the 1960s and 1970s is well known and need not be detailed here, but several aspects of that decline bear mention. Perhaps most important to an assessment of possible causes is the remarkable pervasiveness of the decline. The drop in test scores took place among many different types of students, in many subject areas, on diverse tests, in all parts of the nation, and in Catholic as well as public schools. 2/ Indeed,

---

2. **The achievement decline among private schools evident in nationally representative data largely reflects the drop in scores of students in Catholic schools; the data are insufficient to gauge separately the trends in non-Catholic private schools. See Donald Rock, Ruth B. Eckstrom, Margaret E. Goertz, Thomas L. Hilton, and Judith Pollack, *Factors Associated With Decline of Test Scores of High School Seniors, 1972 to 1980* (Washington, D.C.: Center for Statistics, Department of Education, 1985), Chapter 5 and Appendix D. This distinction between Catholic and other private schools was not noted in *Trends in Educational Achievement*, the report from which this chapter is adapted.**

data on test scores from Canada, though limited, suggest that somewhat similar trends appeared there as well. 3/ Available data do not pinpoint the onset of this decline precisely but suggest that it began in all affected age groups within a short period during the mid-1960s.

Though pervasive, the achievement decline showed substantial variations, and these variations--when they occur consistently in numerous tests--also shed light on possible causes. One of the most important of these differences is that the decline was greater among older students. The decline lasted longer in the higher grades; in addition, limited evidence suggests that scores dropped more rapidly on tests administered to older students, at least during the early years of declining scores. Thus, the tests that have received the greatest attention and that have shaped many observers' impressions of achievement trends--tests administered to high school students--generally showed the greatest drops in scores. In contrast, tests administered in the first three grades showed little or no decline, and those administered in the middle grades tended to show moderate declines.

Another, particularly distressing, variation in the data is that higher-order skills (that is, skills such as reasoning and problem-solving), which showed particularly severe weaknesses throughout the period considered, deteriorated more markedly in some instances than did the most basic skills (such as factual knowledge, literal decoding of written text, and mastery of computational algorithms). The National Assessment of Educational Progress, for example, found somewhat greater drops in performance in higher-order skills in both mathematics and reading.

The greater severity of the decline in scores in the upper grades might also be an indication of the sharper deterioration of higher-order skills, because the material included in tests administered in higher grades is progressively more complex. Indeed, the virtual absence of a decline in scores in the first three grades might partly reflect the emphasis on basic skills in tests administered in those grades. It is important to note, however, that the particularly severe problems with higher-order skills are also apparent even in the case of relatively simple material, including some taught in the elementary and junior-high grades. The National Assessment of Educational Progress, for example, found that large numbers of students

---

3. The primary source of data pertaining to Canadian students is from an adaptation of the Iowa Tests of Basic Skills administered to Canadian students through grade 8 in 1966, 1973, and 1980. See *Canadian Tests of Basic Skills: Manual for Administrators, Supervisors, and Counselors, Levels 5-18, Forms 5 and 6* (Scarborough, Ontario: Nelson Canada, 1984), p. 80; also, Thomas Schweitzer, Economic Council of Canada, personal communication, February 18, 1987.

are unable to apply basic arithmetic algorithms to the solution of simple word problems.

The Upturn in Test Scores

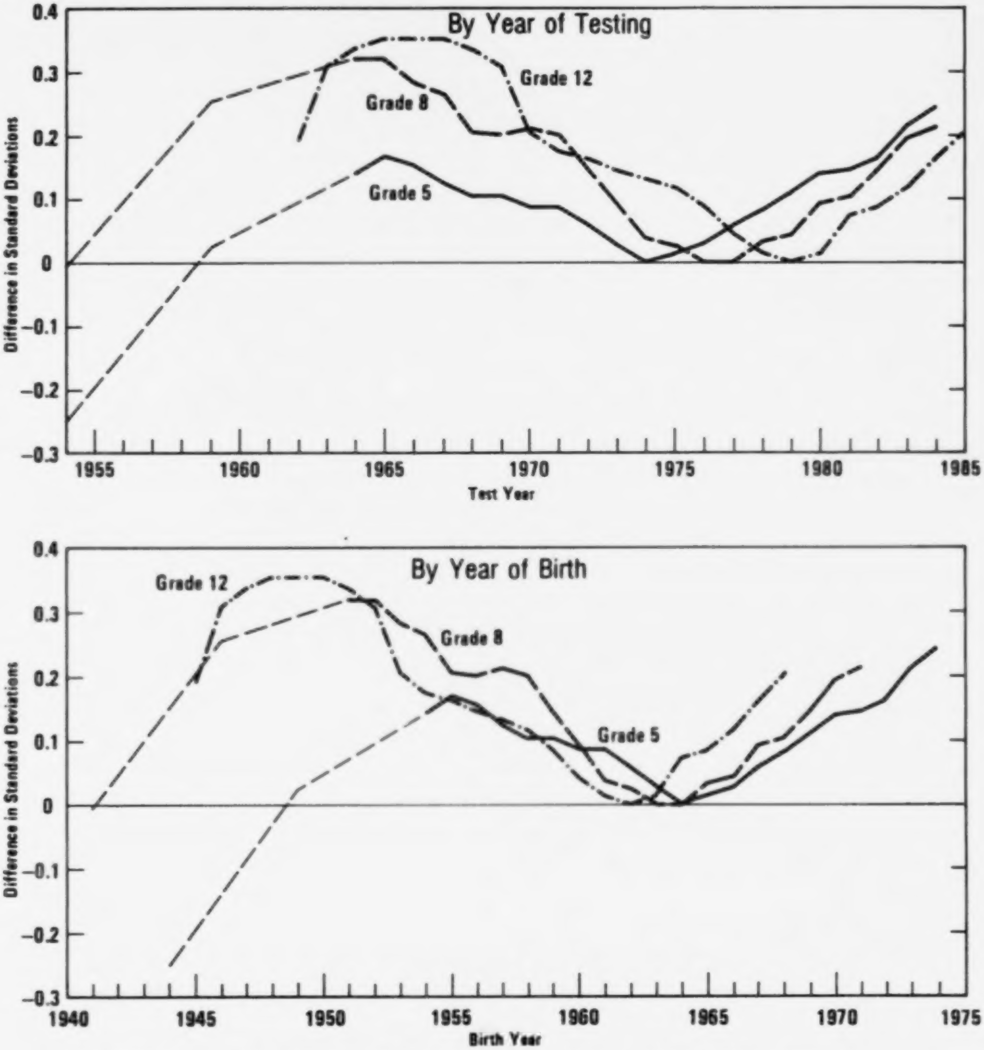
The current debate about education, while still shaped largely by the decline in test scores, has been altered recently by a growing awareness of favorable trends in achievement. It is now generally recognized that a widespread rise in test scores followed immediately on the heels of the decline and has been under way for some time. The characteristics of that upturn, however, are less well recognized. In particular, because of the greater attention afforded to tests administered at the high school level, such as the Scholastic Aptitude Test (SAT), many analysts have mistakenly believed that the rise in scores began within the past few years. In fact, the upturn was apparent in certain grades as early as the mid-1970s.

The upturn, like the preceding decline, was not uniform, and again variations in the trends hold keys to understanding their possible causes. Particularly important are differences among age groups. The decline in test scores ended--and the subsequent rise in scores began--first in the lower grades and later in the higher grades. The upturn first became apparent in test scores of students in the middle elementary grades in the mid-1970s. For example, in the Iowa state assessments--in some respects the best available data on trends in elementary and secondary achievement, although not representative of the nation as a whole--scores of fifth-grade students began climbing in 1975. The upturn then moved into the higher grades at a rate of roughly one grade per year, reaching the senior high school grades around the end of that decade. The end of the achievement decline and the onset of the following rise thus appear to constitute a "cohort effect"--a change that occurs in one or a few birth cohorts and therefore appears in different age groups as the affected cohorts grow older. This reversal in the trends occurred on most tests within a few years of the birth cohorts of 1962 and 1963 and moved up through the grades as those cohorts passed through school. (This pattern is clearest in the Iowa state data; see Figure 2.) Subsequent birth cohorts have typically scored progressively higher.

The upturn in scores in the lower grades has to date been larger than that in the upper grades. By some measures, the rise in achievement in the elementary grades has more than fully overcome the decline, so that scores are now at their highest point on record--a span of up to three decades. In contrast, scores on some tests administered in the higher grades remain considerably below their pre-decline high point. The greater improvement

Figure 2.

Iowa Average Test Scores, Grades 5, 8, and 12, Differences from Post-1964 Low Point



SOURCES: Congressional Budget Office calculations based on "Iowa Basic Skills Testing Programs, Achievement Trends in Iowa: 1956-1985" (Iowa Testing Programs, unpublished and undated material); A.N. Hieronymus, E.F. Lindquist, and H.D. Hoover, *Iowa Tests of Basic Skills: Manual for School Administrators*(Chicago: Riverside, 1982); "Mean ITED Test Scores by Grade and Subtest for the State of Iowa, 1962 to Present" (Iowa Testing Programs, unpublished and undated tabulations); and Robert Forsyth, Iowa Testing Programs, personal communication, August 1984.

NOTE: Dashed lines (– – –) indicate data available only at five-year intervals.

in the lower grades apparently has resulted largely from the longer duration of the rise in the lower grades--that is, the larger number of higher-performing birth cohorts who have so far reached the lower grades. The limited available data suggest that the annual rate of improvement has been roughly comparable in different grades (see Figure 2).

Variations in Trends Among Types of Students and Schools

Achievement trends have also varied among different groups of students. One of the most consistent trends of the past decade has been the gains of black students relative to nonminority students--a pattern that appears without serious exception on every test identified in this study in which separate data for black students are available. Although this pattern results in part from the more rapid deterioration of scores among nonminority students during the last years of the decline, much of the relative gain of black students is real, in that it reflects greater subsequent improvement in their performance than has been shown by nonminority students. The gap in average scores between black and nonminority students, however, remains large on most tests. Hispanic students also appear to have gained relative to nonminority students, although the data pertaining to Hispanic students are less clear-cut.

Because various types of schools are influenced by different educational practices and social trends, information about achievement trends in different types of schools also has an important bearing on explanations of the trends. It is therefore striking that test scores declined among students in Catholic schools in the United States and Canadian schools as well. In contrast, the existing data, though very sparse, suggest that trends in two other categories of schools--those with high concentrations of minority students and those located in disadvantaged urban communities--have diverged markedly from national trends in recent years. Schools in both categories appear to be gaining appreciably relative to the national average.

The Average Level of Performance on Tests

Despite the recent rise in test scores, the average performance among certain groups and, in some instances, nationwide remains distressingly low. Recent National Assessments of Educational Progress (NAEP) in reading, writing, mathematics, and literacy are rife with illustrations of important skills that large segments of the student population are failing to master. These deficiencies are particularly clear in the assessments of high school students and young adults.

The National Assessments of mathematics, for example, indicate that many students are failing to master even fairly rudimentary skills, particularly when they must reason for themselves what skills to apply rather than simply use a specified arithmetic algorithm. Among 17-year-olds still enrolled in school, only 50 percent to 60 percent (depending on the year of the assessment) were able to solve simple problems involving percentages. (An example is the question: "A hockey team won 5 of its 20 games. What percent of the games did it win?") The proportion able to calculate the cost of electricity per kilowatt hour, given a highly simplified electrical bill, varied from 5 percent to 12 percent, again depending on the year. 4/

The National Assessment of literacy conducted in 1986 revealed striking deficiencies in the ability of young adults (ages 21-25) to use written text in a variety of ways. 5/ Less than 40 percent, for example, could synthesize the main argument of a lengthy newspaper article.6/ Roughly 60 percent could extract information from a bar graph, use a chart to pick an appropriate grade of sandpaper, or follow directions using a street map. Given the disturbing level of performance in the mathematics assessments, it is not surprising that some items in the literacy assessment that entailed the use of arithmetic also revealed serious deficiencies. One question presented a simple menu and asked respondents to answer two questions: how much change they would get from a given amount of money

---

4. National Assessment of Educational Progress, *Changes in Mathematical Achievement, 1973-1978* (Denver: NAEP/Education Commission of the States, 1979). These data, which reflect tests administered in both 1973 and 1978, are among the most recent nationally representative data about the mathematics achievement of 17-year-old students. Although current mathematics achievement is probably appreciably higher than that of 1978, it is not likely to be dramatically higher than that of 1973, which was roughly six or seven years before the end of the decline in that age group.

5. The NAEP literacy assessment differed from that of reading in three important respects: the literacy assessment considered a far broader range of skills (including, for example, the ability to apply rudimentary arithmetic operations in solving problems presented in written text); it tested older youths (ages 21-25, rather than ages 9, 13, and 17); and it included in the sample youths who had dropped out of school.

6. The proportions of tested individuals noted here as showing a given skill are only approximate. In contrast to many of the earlier National Assessments, the literacy results were not reported in terms of the proportion responding correctly to specific test items. Rather, the proportion performing at a given level of proficiency was reported, along with one or two items indicative of the skills required to demonstrate that level of proficiency. The proportion responding correctly to one of the illustrative test items would generally be slightly different from the proportion showing that level of proficiency, based on all relevant items.

if they ordered two specific items, and how much would be required for a 10 percent tip. Only about 40 percent correctly answered both questions.

IMPORTANT GAPS AND INCONSISTENCIES IN THE ACHIEVEMENT DATA

Examination of a broad array of achievement tests adds considerably to the information that can be obtained from any single test, even if that test yields data of particularly high quality. Yet examining the available test data also reveals the limits of what is currently known about educational achievement. A number of important questions are simply not adequately addressed by available data, and some conclusions that appear straightforward in a single source of achievement data are shown to be questionable when many sources are considered. These gaps and inconsistencies in the data are important not only for understanding the condition of education, but also for explaining recent trends; some of the common explanations are based on aspects of the recent trends that are striking in the results of one or two tests but fail to appear--or are contradicted--in the results of others.

The inconsistencies in the existing test data affect even some of the most fundamental conclusions about recent trends. For example, the size of the decline differed substantially among tests. Tests have also offered dramatically dissimilar pictures of relative trends among different subject areas--an important pattern for explaining the trends, because many explanations are based on factors that would affect some subjects more than others. Regional differences in trends have also varied among tests: the National Assessments have tended to show more favorable trends in the South, which is by some measures the lowest-scoring region, than elsewhere. On the other hand, the only other nationally representative study of regional disparities in trends indicated that declines in scores among high school seniors in the South ranged from being comparable to those elsewhere in one subject to being far worse in another. 7/

---

7. These latter results reflect a comparison of the National Longitudinal Study of the High School Seniors Class of 1972 and the High School and Beyond study. See Rock and others, *Factors Associated With Decline of Test Scores*, Appendix D. Rock used standard Census definitions of the regions, while the National Assessment included in other regions several states that the Census classifies as part of the South. The difference between the results of the two studies is so large, however, that it is very unlikely that this discrepancy in definitions could account for it.

It is also uncertain from the available data whether the trends in test scores vary consistently among achievement subgroups--that is, among groups differing in their initial levels of achievement. Relatively favorable trends among low-achieving students appeared clearly in the NAEP and have figured prominently in some explanations of recent trends. When one considers a variety of tests, however, the information on relative trends among achievement subgroups appears to be a welter of inconsistent findings and disparate definitions of groups. Moreover, comparison of trends among achievement subgroups is hindered by a number of serious technical obstacles. The use of alternative (and equally defensible) methods of scaling and reporting test scores, for example, can fundamentally alter the conclusions one reaches, and the published data are insufficient to sort through the resulting confusion.

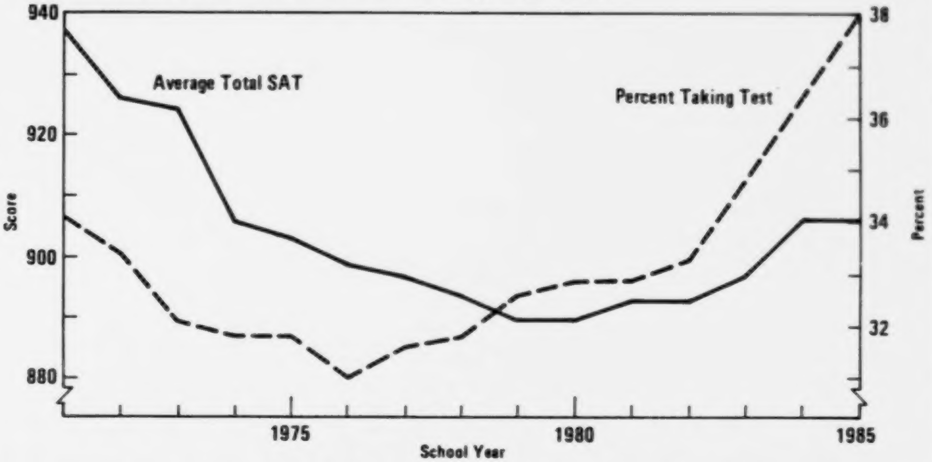
Also unanswered is the question of whether the recent rise in scores is beginning to falter. The data offer little reason to doubt that scores in the higher grades will continue rising for several years as the cohorts that recently produced gains in the lower grades progress through school, just as earlier gains in the lower grades were echoed later in the higher grades. Any number of factors could deflect those trends--either augmenting the gains or lessening them--but the data as yet do not indicate such a change (see box on facing page). In contrast, some achievement tests have shown stable scores in the early grades during the past few years, while other tests have shown continuing gains. Only the accumulation of additional information over the next few years will clarify whether progress in the lower grades has indeed ceased for the time being and, if so, whether that stagnation will be duplicated in the higher grades as the affected cohorts progress through school.

HAVE HIGH SCHOOL TEST SCORES STOPPED RISING?

In the 1985 school year, average SAT scores remained at the level of the previous year, seemingly ending an erratic but appreciable rise that had been under way for half a decade. Some analysts quickly seized on this as evidence that the rise of achievement at the senior high school level had stagnated, even though no other major source of data suggests that scores have stopped rising in those grades.

A closer look at the SAT scores, however, shows that the current stability of average scores probably does not indicate that student performance has become stagnant. Beginning in the mid-1970s, the share of high school graduates taking the SAT grew sharply, from 31 percent in 1976 to 38 percent in 1985. Just as a similar growth in the test-taking group exacerbated the SAT decline in the 1960s, the current increase probably impeded the rise in SAT scores substantially. That is, as the pool of test-takers grows, it generally also becomes less selective, and the addition of lower-scoring students depresses average scores. If the proportion of graduates taking the test had remained constant, SAT scores would have been a better gauge of changes in student performance--but they also probably would have risen more, and 1985 scores might well have been higher than those of 1984.

The SAT: Average Scores
and the Percent of Graduates Taking the Test



SOURCES: Congressional Budget Office calculations based on The College Entrance Examination Board, *National College-Bound Seniors* (New York: The College Board, various years); Office of Educational Research and Improvement, *Digest of Education Statistics, 1987* (Washington, D.C.: Department of Education, 1987); and Office of Educational Research and Improvement, unpublished data.

CHAPTER III

APPROACHES TO EXPLAINING ACHIEVEMENT TRENDS

One can easily enough devise plausible explanations of recent trends in educational achievement. The quantity and diversity of explanations that have been advanced to date give ample evidence of that. Indeed, many of the common explanations seem so persuasive that they have been subjected to relatively little scrutiny, even when they provide the rationale for formulating policy initiatives.

Yet there are many reasons to be cautious in ascribing trends to causes. Some of the common and influential explanations turn out on closer examination to be wrong; others cannot be tested with existing data. Still others appear plausible but could account for only a very small share or some particular aspect of the total movement of average test scores. Moreover, even when a factor could plausibly have contributed to the trends, the evidence typically affords much less certainty about its effects than many observers had expected.

Some attempts to explain the trends have gone amiss because they failed to distinguish between a factor's contributions to the trends and its effects on achievement more generally, and the conclusions of this analysis could likewise be misinterpreted if this distinction is not borne in mind. If this study examined only the factors' effects on achievement more generally, the methods used would be simpler, and the conclusions would in some instances be significantly different.

One approach to explaining the achievement trends is to evaluate the evidence pertaining to individual causal factors. Does the evidence suggest that changes in textbooks, for example, indeed contributed to the trends? To what aspects of the trends might they have contributed, and how big might their effects have been? By considering many diverse factors, one can gradually develop from these pieces a general view of the trends' causes. This factor-by-factor approach has characterized much of the debate to date. But many of the assessments have been incomplete, and few analysts have tried to piece the various conclusions together into a general view of the trends' possible causes.

A complementary approach, also used in this analysis, starts with the whole rather than with specific causal factors. Given the broad patterns of the achievement trends, what can one infer about likely causes? For example, one can reach different sorts of explanations on the basis of variation or lack of variation in trends among private and public schools, types of communities, age groups, students of different ability, and so on. This approach has been taken less often, perhaps because information about the broad outlines of recent trends was limited until quite recently.

EVALUATING EVIDENCE ABOUT SPECIFIC FACTORS

The first of these two methods--analyzing the evidence pertaining to one causal factor at a time--appears straightforward. In practice, however, the analyst must bear in mind a number of considerations.

Criteria for Evaluating the Effect of Specific Factors. Ideally, two criteria should be applied in evaluating the contributions of specific factors to achievement trends. The first criterion is whether a factor shows any relationship with achievement in cross-sectional studies--that is, whether an association exists between that factor and achievement levels at any given time. For example, among this year's high school seniors, do those who do more homework score better on achievement tests, all other things being equal? The second criterion--called temporal consistency here--is whether changes in an explanatory factor over time are consistent with trends in achievement. For example, have changes in the amount of homework done by typical students paralleled changes in average test scores?

Affirmative evidence about both cross-sectional relationships and temporal consistency is usually required to support a proposed explanation of the achievement trends; negative evidence about either criterion can be sufficient to refute it. Some key misconceptions about recent trends in achievement appear to have arisen because one or the other of these two criteria was paid too little heed.

No matter how strong the cross-sectional evidence pertaining to a given factor, the analyst must show temporal consistency in order to link it to specific trends in test scores. A factor that is shown by cross-sectional research to be a powerful influence on achievement in general can still be temporally inconsistent with specific trends in achievement and therefore incapable of having directly contributed to them. The importance of temporal consistency is perhaps clearest in cases where a factor of interest showed no change during the relevant period. If the amount of television viewed, for example, did not change at all during the period of the trends in achievement being examined, one could conclude even without cross-sectional data that, whatever the effects of television viewing on achievement in general, the specific trends in test scores cannot be attributed to changes in viewing. By the same logic, finding that a certain factor could not have contributed to these trends because it was temporally inconsistent with them need not imply that it has no effects on achievement more generally or that it will not influence future trends in scores.

A lack of temporal consistency is not a problem, however, in the case of many common explanations of the achievement trends; in fact, they were first suggested precisely because they do show temporal consistency with some aspect of recent achievement trends--often congruity with trends in scores on a single test. The problem with many of these explanations is that temporal consistency alone provides little basis for concluding that a factor contributed to the trends in test scores. Innumerable factors can be found that show trends over time that are reasonably consistent with some particular aspect of trends in test scores, and yet many of these factors had no bearing on the achievement trends. To link these factors to the achievement trends, one needs some basis for judging them capable of influencing test scores. In some instances, the link may be so obvious that analysts feel no need to substantiate it. In most cases, however, cross-sectional evidence is required to establish the link.

Obstacles to Evaluating Specific Explanations of the Achievement Trends. In their efforts to assess cross-sectional evidence and temporal consistency, researchers encounter a number of important obstacles.

In many instances, inadequacies of the existing data impede--or even preclude--an assessment of cross-sectional evidence or temporal consistency. Data about many potential causal factors are of poor quality or lacking altogether. Moreover, even when the potential explanatory factors themselves have been reasonably well measured, cross-sectional information may be so weak in other ways that only tentative conclusions--or no conclusions at all--about the factors' possible effects are warranted.

A particularly common problem in the research reviewed here is the omission or inadequate treatment of other variables--called confounded variables--that are associated with both the factors of interest and achievement and that might be responsible for the associations between them. For example, studies showing a relationship between the amount of homework and students' test scores tell little about the value of homework itself unless the studies isolate the impact of other characteristics of students who do a lot of homework and also score well on tests. Such factors might include the students' aptitude, previous achievement, and motivation. Similarly, many studies of the relationship between class size and achievement fail to take into account decisions in some schools to assign low-achieving students to small classes in an effort to improve their performance.[1] Because a beneficial effect of smaller class sizes might be masked by the lower potential of students assigned to the smaller classes, such studies do not provide a good assessment of the independent effect of class size.

Even when the quality of existing data is not an obstacle, gauging temporal consistency may be complicated by the long duration of schooling. For example, in looking at a test that is administered to students after 11 years of schooling, such as the SAT, one has to decide which point during those 11 years to align with potential explanatory factors. Some analysts have searched for factors that were temporally consistent with the scores themselves, such as changes in various aspects of high school education. An alternative view is that the early years of schooling are important determinants of achievement in later grades. One analyst, for example, arguing that the early years of schooling are major determinants of reading ability, attributed trends in SAT scores to changes in the teaching of reading in primary grades a decade before each cohort took the SAT. [2]

Efforts to assess temporal consistency are also made more difficult by the complexity of the achievement trends themselves. When one considers a wide array of achievement tests, it becomes apparent that factors that have been singled out for attention because of their consistency with a single aspect of the achievement trends are inconsistent with other aspects. In some instances, the inconsistencies that are revealed make an explanation appear implausible altogether, but in other cases, they help clarify what the specific effects of a factor could have been. For example, some factors that have been offered as explanations are temporally consistent with test score trends in the higher grades but inconsistent with those in the elementary grades. Such factors could not have initiated trends that the relevant cohorts first exhibited when they were in the earlier grades, though they might have helped increase the severity of those trends in the higher grades.

---

1. Some of these studies also fail to consider the association between class size and the size and location of schools, which are in turn related to achievement. For a discussion of all these omissions, see Stephen N. Simpson, "Comments on 'Meta-Analysis of Research on Class Size and Achievement,'" *Educational Evaluation and Policy Analysis*, vol. 2 (May-June 1980), pp. 81-83.

2. See, for example, Jeanne S. Chall, "Literacy: Trends and Explanations," *Educational Researcher*, vol. 12 (November 1983), pp. 3-8.

A final impediment to reaching firm conclusions is that available data generally show only that certain factors are associated with achievement and usually cannot demonstrate that those factors actually caused the trends. For example, the achievement decline among high school seniors in the 1970s was associated with a drop in the proportion of high school students enrolled in academic programs. This change might have contributed to the decline in seniors' test scores. Alternatively, both this change and the decline in test scores might have been the effects of still other factors, such as a drop in students' motivation or in their achievement at earlier grades. Indeed, both explanations could be correct. The available data are inadequate to disprove either of these competing interpretations of the observed association.

INFERRING CAUSES FROM GENERAL PATTERNS IN THE TEST SCORE DATA

The second approach to assessing the trends' causes--inferring them from the broad patterns in the achievement data--leads to very general conclusions. It might suggest, for example, that societal factors (such as demographic, cultural, and other noneducational factors) contributed to a certain aspect of the trends in achievement but give few clues about which societal factors might have been germane. The conclusions it yields are also more inferential and arguable than are those based on the assessment of individual factors. Nonetheless, this approach yields some of the most important conclusions about the trends' causes.

This alternative approach requires that one go beyond information from a single or even a few tests to discern the common threads and important divergences among various sources of test score data. One example, discussed in more detail in the following chapter, is the consistency or variation of achievement trends among diverse settings and subject areas. Despite important gaps, the available achievement data are abundant enough to make this approach possible.

To draw inferences of this sort, one needs to look not only for achievement patterns consistent with a given type of explanation, but also for patterns that are inconsistent with the alternative explanations. For example, consider the hypothesis that societal factors contributed to the trends in test scores. That hypothesis would gain support, not only from patterns in the data that could plausibly reflect societal factors, but also from patterns that are difficult to explain in terms of educational factors.

CHAPTER IV

THE POSSIBLE CAUSES OF THE ACHIEVEMENT TRENDS

Many explanations of recent trends in test scores attribute them to one or a few factors. Moreover, many analysts have focused on a single type of possible cause. One group of analysts--probably the largest and certainly the most influential--holds that the decline and subsequent upturn of test scores were largely the result of educational factors, many of which can be directly affected by explicit changes in educational policy. Another group places more of the responsibility on noneducational factors, some of which (demographic trends, for example) it sees as neutral and uncontrollable, and others of which (such as drug abuse) it regards as value-laden and alterable.

Although these views appeal to common sense and contain elements of truth, they are largely wrong. The available evidence suggests that the trends resulted from the confluence of many causes, both educational and noneducational, not from one or a few powerful factors. The individual contributions of those causes, when they can be estimated, appear to have ranged from very small to modest. In addition, many of the factors that have been cited with particular frequency turn out on closer examination to have played no role at all, and the importance of other factors cannot even be tested for want of appropriate data.

THREE GROUPS OF CONTRIBUTING FACTORS

The factors that plausibly could have contributed to the trends are extremely diverse. These factors can be organized into three broad categories:

- Modifications of educational policy, conditions, and practice;
- Changes in the selection of students to be tested--commonly called selection factors; and
- Broad societal and cultural trends.

Educational factors include explicit modifications of educational policy, such as changing criteria for promoting students into subsequent grades, adopting easier textbooks, and changing the range of courses that secondary school students are offered or required to take. The category also includes trends in educational practice that might go beyond those resulting from explicit policy changes, such as changes in the length or frequency of homework assignments, the extent of "teaching to the test," and teachers' expectations of their students. Other changes in the condition of the educational system, such as trends in the amount of experience or educational background of teachers, are also included in this category.

The term "selection factors" here refers to changes in which students from a group of potential test-takers--for example, which children of a given age--are tested. 1/ Selection changes can stem from trends in enrollment, such as changes in the initial enrollment rates of various groups, retention rates (that is, the proportion of students from different groups remaining enrolled until a given age or grade), and the proportion of students from different groups who fall behind the typical grade level for their age. 2/ Testing policy also can affect selection--for example, by determining which out-of-grade students, or which children with certain handicaps or with limited proficiency in English, are tested. Finally, one important aspect of selection--called self-selection--reflects students' decisions to take optional tests. Self-selection is primarily relevant to college admissions tests, such as the Scholastic Aptitude Test (SAT) and tests of the American College Testing Program (ACT).

The category of societal factors comprises all factors that are neither educational nor selection-related. It includes family composition, participation of mothers in the labor force, cultural factors such as students' attitudes toward educational success and career options, the ethnic composition of the student population, and environmental factors such as children's exposure to toxic substances.

The meaning of a change in test scores depends in part on which of these three categories is responsible for it. Test score trends attributable

---

1. The use of "selection factors" is much more specific than the more common but vaguer concept of "compositional changes." The latter concept includes all changes in the composition of the test-taking groups, regardless of whether they stem from selection or from changes in the makeup of the cohort from which the test-taking group is drawn. For example, a change in the ethnic composition of the test-taking group is a matter of selection if it stems from a change in the dropout rate among black students, but not if it reflects trends in the makeup of the school-age population as a whole. As explained below, the significance of resulting changes in test scores may hinge on this distinction.

2. The proportion of students falling behind the typical grade affects the mix of students tested, because routine testing is commonly linked to grade levels rather than age.

to educational factors represent clear-cut changes in student performance; they reflect changes in the success with which specific skills are imparted to students, not changes in the characteristics of students entering school or selected for testing. In contrast, trends in test scores attributable to selection factors should rarely be construed as real changes in achievement or in the success of instruction. Rather, they are simply artifacts of changes in which students are chosen--or choose themselves--for testing.

To clarify this distinction, consider a high school that institutes a new program that reduces its dropout rate by half. If this change has no effect on the performance of students who would have remained in school in the absence of the new program, one would expect the school's average scores to decline because of the lower scores of students who otherwise would have dropped out. This apparent deterioration of scores, however, would indicate nothing other than the changed selection of students from the population of youth in that school's attendance area. Indeed, if the achievement of the students who were prevented from dropping out rose as a result of their remaining in school, the decline in average scores would actually be masking a real increase in the achievement level of the cohort as a whole.

Trends in test scores attributable to societal changes fall in between. Their meaning varies depending on the question at issue and the particular societal factors involved. A decline in scores attributable to a pervasive drop in students' motivation, for example, would generally be seen as a true decline in achievement. In contrast, the interpretation of a decline in scores stemming from changes in the ethnic composition of the entire school-age population is more ambiguous, assuming that scores within each ethnic group remain unchanged. If the question of interest is the achievement level of the cohort as a whole, such a decline represents a true change in performance. But if the concern is with the effectiveness of an educational system, many analysts would see such a decline as similar to a selection change, since it would not signify a deterioration of the educational performance of students from any given ethnic group.

THE EFFECTS OF EDUCATIONAL, SELECTION, AND SOCIETAL FACTORS

The evidence suggests that educational, selection, and societal factors all contributed, though in different ways, to the decline in test scores that occurred in the 1960s and 1970s.

The contributions of both educational and societal factors to the decline appear to have been considerable, and numerous factors in each group played a role. The separate effects of individual factors, however, were in most instances apparently small. The data are insufficient to indicate the relative importance of the two categories. The known contributions of selection factors to the decline, on the other hand, were limited to scores on certain optional tests taken by high school students--in particular, the SAT and ACT. Nonetheless, because the college admissions tests that were affected are among the tests that have commanded the greatest attention, selection factors have had a major effect on the public's perception of the decline in achievement.

It appears that both educational and societal factors contributed significantly to the subsequent rise in test scores as well. In contrast, insofar as they have been measured, selection changes have not contributed materially to the rise in scores and may have impeded it in some cases.

The Effects of Specific Factors

As noted in Chapter III, one approach to evaluating the origins of recent achievement trends is to examine the evidence relevant to specific factors that have been suggested as possible causes. More than two dozen such factors have been evaluated for this paper; a discussion of the evidence pertaining to each can be found in the Appendix. Many of the factors included here are frequently cited and have been particularly influential in shaping public perceptions about the causes of the achievement trends. Several factors that only rarely have been noted in this context are also included because their impact on test scores could be significant. A great many factors have been suggested as having contributed to recent achievement trends, however, and the subset discussed here is necessarily incomplete. The omission of other factors from this study does not imply that they were unimportant.

The factors considered here can be grouped into three categories: those that are plausible causes of some aspect of the trends, those that probably did not contribute appreciably, and those that cannot be assessed because there is insufficient evidence.

The factors that remain as plausible causes when systematic evidence is considered include educational, societal, and selection factors (see Table 1 and the Appendix). Although the relative importance of educational and societal causes cannot be determined, the contribution of the latter was clearly considerable. Indeed, two factors that made particularly substantial contributions to the decline were societal: changes in the ethnic composition of the school-age population, and trends in family size (that is, the number of children per family and average birth order). 3/

A number of commonly cited educational factors could have contributed to certain aspects of the trends. A weakening of course content in the secondary grades might have contributed to the achievement decline and might help explain the greater severity of the drop among older children. Changes in the amount of homework might have contributed to both the decline and the subsequent upturn, at least among high school students. A drop in the proportion of teachers with little experience might have aided the upturn in scores, although an earlier decline in the average experience of teachers was probably unrelated to the decline in test scores.

Educational factors might also have contributed to the relative gains of minority students--that is, to the narrowing of the gap between their scores and those of nonminority students. Chapter 1 (the federally funded compensatory education program, formerly Title I) could account for some of the relative gains of both black and Hispanic students, but its contribution to this specific pattern is limited by the large proportion of nonminority students in the program and by the relatively small share of the student body that participates in the program. 4/ Desegregation also might have contributed to the relative gains of black students, but it could not have influenced the gains of Hispanics, for they did not become less segregated.

Societal factors that might have contributed to the trends are diverse and include some that have been prominent in the debate about educational achievement and others that have received little attention in this regard. Changes in the ethnic composition of the student body appear to have contributed significantly to the decline in scores but probably impeded the subsequent rise. Changes in family size brought about by the baby boom and baby bust, which have been the focus of considerable attention, probably contributed to both the decline and the upturn. Trends in the use of alcohol

---

3. Birth order refers to the sequence of births in a family; a first-born child has an order of one, a second-born, two, and so on. The average birth order of a cohort is simply the average order of all children born in that year.

4. Chapter 1 could account for far less of the relative gains of minority students in the higher grades because of the much smaller number of older students served by the program and because the program's effects on test scores apparently largely erode after several years.

TABLE 1. THE EFFECTS OF VARIOUS EDUCATIONAL, SOCIETAL, AND SELECTION-RELATED FACTORS ON RECENT TRENDS IN TEST SCORES

Factor	Effect
Factors that Could Have Contributed	
Educational	
Teachers' Experience	May have contributed to upturn in scores but not to decline; effect cannot be quantified
Coursework	Change in content rather than number of courses probably contributed to the decline; cannot be quantified
Textbook Characteristics	Evidence limited to a few subjects; cannot be quantified a/
Homework	Possible small contribution to both decline and upturn
Title I/Chapter 1 Compensatory Education	Possible modest contribution to the relative gains of black and Hispanic students (compared with nonminority students), but only in the early grades; possible slight contribution to the relative gains of younger students
Societal	
Desegregation	Could account for a modest share of the relative gains of black students but not for the gains of Hispanic students
Ethnic Composition	Could have contributed one-tenth to one-fifth of the decline but impeded upturn b/

(Continued)

TABLE 1. (Continued)

Factor	Effect
Family Size (Number of children per family, average birth order)	Could have contributed modestly to both decline and upturn; best estimates range from 4 percent to 25 percent but are probably too high b/
Alcohol and Drug Use	Increase in use might have contributed to decline in higher grades; subsequent drop in use might have contributed to rise in scores
Environmental Lead	Reduced exposure to lead might have made a small contribution to the upturn
Selection	
Self-Selection	Contributed appreciably to decline on college admissions tests and might have impeded rise on those tests, but irrelevant to other tests
Factors that Probably Did Not Contribute Significantly	
Educational	
Teachers' Test Scores	Change after 1972 did not contribute to the decline; earlier data are not available
Teachers' Educational Attainment	Not temporally consistent with decline in test scores
State Graduation Requirements	No direct contribution to the decline after 1974; earlier data are not available
Minimum-Competency Testing	Did not help initiate the upturn; other effects are uncertain

(Continued)

TABLE 1. (Continued)

Factor	Effect
Head Start	No appreciable contribution to relative gains of black or Hispanic students (compared with nonminority students) after third grade; inconsequential contribution to relative gains of youngest students
Societal	
Single-Parent Households	Inconsequential contribution to the decline among young children; probably no appreciable effect among older children
Maternal Employment	Inconsistent data about relationship to achievement c/
Television Viewing	Amount of viewing did not parallel achievement trends
Selection	
Retention Changes	Little or no direct contribution after about 1968
Factors About Which There is Insufficient Evidence	
Educational	
Other Characteristics of Teachers (Such as attitudes and morale)	
Local Graduation Requirements	

(Continued)

TABLE 1. (Continued)

Factor	Effect
Grade Inflation	Inflation has been documented, but its effects have not
Demands for Writing d/	
Societal	
Students' Attitudes and Motivation	
Selection	
Other Selection Changes (In the testing of handicapped children, for example)	

SOURCE: Congressional Budget Office.

NOTE: For further explanation and documentation, see Appendix.

a. Evidence in some subject areas indicates no effect.

b. Estimates reflect only part of the relevant period.

c. Cross-sectional evidence about the effects of maternal employment are inconsistent. Because future studies might resolve these inconsistencies, maternal employment could also be placed in the "insufficient evidence" category.

d. Demands for writing might also be placed in the "probably did not contribute" category. Available systematic data do not indicate relevant changes in this factor but are too sparse to yield a firm conclusion.

and other drugs by high school students might have contributed to both the decline and the upturn; like changes in the content of secondary school coursework, this factor might help explain the greater severity of the decline in the higher grades. A pervasive decline in exposure to environmental lead--a factor that has received extensive attention as an influence on children's health and cognitive functioning but that has rarely been mentioned as a possible cause of the achievement trends--might also have contributed to the recent rise in test scores.

It appears that the contributions of these factors were generally less substantial than many observers have thought, ranging from very small to modest. Two factors whose effects can be estimated relatively well--changes in ethnic composition and family size--each can account for at most a fifth to a fourth of the total change in scores during portions of the decline. Although the contributions of some other factors cannot be estimated well, it appears likely that some had effects that were considerably smaller.

The factors whose hypothesized contributions to the trends are not supported by the data include some--both educational and noneducational--that have had broad acceptance and considerable influence in the public debate. For example, despite widespread concern about the effects of declining test scores of teachers, the documented decline occurred too late to have contributed to the deterioration of students' test scores. State graduation requirements have also been the focus of extensive attention but showed no appreciable change during the latter half of the decline in students' scores. (Both of these variables might have played some role, however, during the first half of the decline; the existing data do not extend back far enough to answer that question.)

Minimum-competency testing is another example; whatever its more recent effects on achievement in general--a contentious question that this analysis does not attempt to resolve--its implementation came too late to help initiate the upturn in achievement in the 1970s. A number of common societal hypotheses also fail to weather scrutiny. The rising proportion of students living in single-parent households could have contributed at most an inconsequential share of the overall decline in test scores in the early grades and probably no appreciable share of the much larger decline in the higher grades. Regardless of whether television viewing affects test scores in general, it could not have contributed significantly to the decline in test scores of the 1960s and 1970s, since the amount of viewing did not change consistently with the trends in test scores.

Finally, there is simply not enough systematic evidence to assess the effects of a number of other commonly cited factors. This gap in information is serious, for it affects some factors--such as trends in students' attitudes and motivation, certain characteristics of teachers, and local graduation requirements--that conceivably could have had a substantial impact on test scores.

Additional Inferences About the Causes of the Achievement Trends

Because the results of a factor-by-factor analysis are incomplete and leave some of the change in test scores unexplained, many analysts would like to go beyond it. One way to do so would be to extend that analysis to include additional specific factors. This approach, however, is analogous to building a large house from small bricks; given the apparently small contribution of the factors considered here, the list of factors assessed might have to be expanded substantially to obtain a full explanation of the trends. In addition, because of gaps in the data, many factors would remain unassessed, and the explanation would remain correspondingly incomplete.

An alternative approach, noted in Chapter III, is to examine the general patterns apparent in the achievement data for hints about the trends' causes. Two important inferences suggested by this approach are discussed here.

The Contribution of Noneducational Causes. An important inference to be drawn from the broad patterns in the achievement data is that however important the contributions of educational factors, societal factors also probably contributed substantially to the trends in test scores. This inference corroborates the factor-by-factor analysis reported above. Three aspects of the achievement data point to this conclusion:

- The consistency and near ubiquity of the basic trends;
- The cohort effect shown by the timing of the end of the decline and the onset of the subsequent upturn; and
- The parallels in timing between the achievement trends and changes in certain characteristics of American youth.

The strength of this conclusion rests on two judgments: how likely it is that noneducational influences could have produced these particular aspects of the achievement trends, and how difficult it would be to explain them solely in terms of educational factors. Educational factors could have exerted a powerful influence on achievement trends and still be insufficient to account for these particular patterns.

Numerous societal factors would probably affect many students in a broad variety of settings and thus could contribute to the pervasiveness of trends evident in the data on test scores. Any effects of changes in family configuration accompanying the baby boom and the subsequent baby bust, for example, would have been felt throughout the nation, though not equally in all areas. The effects of changes in the ethnic composition of the school-age population would also be widespread, though with local variations. Some of the less measurable societal factors that have been suggested as causes of the achievement trends might also have affected diverse students in many, highly dissimilar settings. These factors include students' increased sense of alienation and their lessened motivation to achieve.

Regardless of the magnitude of their contribution to the achievement trends, educational factors alone seem far less likely than societal factors to have produced such a striking consistency of trends in diverse settings. The highly decentralized nature of the American educational system--in which decisions about educational policy are made by 50 state education agencies, legislatures, and governors, as well as more than 15,000 local education agencies--would tend to lessen the uniformity of achievement trends attributable to educational practices. Despite this decentralization, similar educational changes sometimes do occur in many jurisdictions, and educational factors therefore cannot be ruled out as possible contributors to pervasive trends in achievement. But it is difficult to imagine educational changes sufficiently ubiquitous, extensive, and uniform in timing to have caused by themselves achievement trends as pervasive as those that have occurred over the past 20 years. The evidence of similar trends in Catholic schools and in Canada makes a purely educational explanation even less likely, because Catholic schools are substantially--and Canadian schools entirely--independent of the governance structures that determine policy in American public schools.

Moreover, most educational changes would probably not produce the observed similarity in test score trends among subject areas, types of students, and types of schools. For example, some people have pointed to changes in reading or mathematics curricula as having contributed to the achievement decline, and there is evidence that such changes might indeed have played a role. The principal effects of such changes would presumably be found in those specific subject areas, however, and therefore they would be insufficient to explain the comparable--indeed, in some instances, larger--declines in other subject areas, such as social studies and natural sciences. Similarly, the effects of many of the educational changes suggested as causes of the decline would have been largely limited to certain groups of students--for example, those in specific grades or particular tracks. In contrast, various societal trends, such as demographic changes and shifts in students' attitudes toward schooling, would be quite likely to affect performance more generally.

The cohort pattern shown by the timing of the end of the decline also suggests the importance of noneducational factors. In order to account for this pattern, some of the major influences on test scores must have been experienced by a very large number of children in diverse settings no later than the age of nine, and this set of factors must have acted on cohorts of children, not on students of various ages in school in any one year. Some societal factors, such as changes in ethnic composition, exposure to certain environmental toxins, and perhaps certain aspects of family composition, would fit this pattern.

In contrast, a cohort pattern--and, in particular, the cohort pattern shown by test scores in recent years--is more difficult to explain solely in terms of educational changes. Although several of the commonly cited educational factors might have contributed to the cohort pattern, they appear insufficient, even as a group, to explain it. For example, educational changes at the high school level, such as trends in the tracking of students into academic and nonacademic programs, could have contributed to the cohort pattern by delaying the onset of the upturn in the higher grades, but they cannot explain the existence of the cohort pattern in the upper elementary and junior high school grades. 5/

Finally, the rough parallels in timing between trends in test scores and changes in a variety of other characteristics of American youth suggest that noneducational causes were significant. The suicide, homicide, and arrest rates among white male adolescents and young adults, for example, soared during the years of the test score decline, and the rates among females increased appreciably; more recently those rates have stabilized or declined. The rate of births to unmarried white adolescents also climbed sharply

---

5. Educational changes could have created the specific cohort pattern shown by test scores if they were implemented in all grades above the third, were undertaken first in the early grades and successively later in higher grades, and were undertaken in the late elementary grades fully a decade before the changes in senior high school test scores. Few of the educational factors suggested as possible causes of the achievement trends, however, meet any of these criteria, let alone all three. Educational changes implemented only in the lower grades could also have produced the cohort effect, but only if two conditions were met: if those changes were broad enough to affect achievement in most subject areas, and if their effects were lasting.

during the period of declining test scores. 6/ The societal and cultural shifts underlying trends of this sort might have contributed to a deterioration of test scores as well.

The Timing of Educational Causes. To the extent that educational factors account for the achievement trends, one can infer from the timing of the trends which period's policies might be responsible.

Because trends shown by tests administered at the high school level have commanded the greatest attention, many analysts have searched among the policies of the late 1960s and the 1970s, when scores in the higher grades were falling, for educational practices that might have had deleterious effects on achievement. Similarly, in searching for causes of the upturn in scores, some analysts have looked at policies first implemented in the very late 1970s and 1980s, when scores began rising in the higher grades.

This view is partly correct; for example, certain educational practices of the late 1960s and the 1970s, such as changes in mathematics texts and a watering down of senior high school course content, could have contributed to the decline. But such a view probably obscures some of the important determinants of the trends. It appears just as reasonable to look to that period for policies that might have contributed to rising test scores as to look there for deleterious influences.

Three factors point to this conclusion: the cumulative nature of achievement, the long duration of schooling, and the cohort pattern shown by test scores in recent years. The cohorts that began their schooling in the late 1960s and 1970s have produced unremitting gains in test scores, and the policies in effect during their early years of schooling might have contributed to that improvement. Because the gains produced by these cohorts were evident very early in their school careers--roughly, by the fourth grade--it is even more reasonable to search among their early educational experiences for contributing factors. In view of these considerations, assuming that policies were detrimental merely because they coincided with trends in senior high school test scores appears unwarranted and misleading.

---

6. Edward A. Wynne and Mary Hess, "Long-Term Trends in Youth Conduct and the Revival of Traditional Value Patterns," *Educational Evaluation and Policy Analysis*, vol. 8, no. 3 (Fall 1986), pp. 294-308. In contrast, the rate of births to unmarried black adolescents fell, though erratically, during that period. See National Center for Health Statistics, *Monthly Vital Statistics Report*, vol. 34, no. 6, Supplement (September 20, 1985), Table 18.

The trends in test scores of minority students further suggest a more complex and cautious appraisal of the policies of the recent past. The relative gains of black students, for example, began at least as early as the cohorts that entered school in the early 1960s and were apparent even at the senior high level by the middle of the 1970s. Moreover, absolute gains in scores appeared earlier among black students than among their nonminority peers. If educational factors caused those trends, those factors might have coexisted with other practices that were depressing the test scores of certain nonminority students.

CHAPTER V

IMPLICATIONS

The continuing debate about the quality of public education in the United States, and the accompanying rush of educational policy initiatives throughout the nation, have heightened the importance of understanding recent trends in test scores. Many of these initiatives have reflected a concern that students' achievement is inadequate, or have been intended as a response to recent trends in achievement.

When the current debate and "reform movement" got under way early in this decade, much less was known about recent trends in test scores and their possible causes. The more comprehensive overview of the trends and their causes provided in this paper and in the previous companion study offer a basis for reexamining earlier assumptions and conclusions as educational policy continues to evolve.

This chapter discusses some of the implications of the recent achievement trends and their causes. It is limited, however, to issues addressed in this paper and in the earlier CBO study. Many equally important issues about educational tests and policy are therefore omitted. For example, the use of fixed cut-off scores on minimum-competency tests as a criterion for high school graduation--a common component of recent educational innovations--has generated considerable research and debate. That controversy is not addressed here because the analyses in these two papers offer little clarification of the issues involved. Similarly, the issue of possible bias in the testing of certain ethnic minorities is not discussed. Despite its great importance, that question is neither critical to understanding the relative trends among ethnic groups discussed here nor illuminated by this analysis.

ASSESSING EDUCATIONAL ACHIEVEMENT

Although many of the basic questions about trends in educational achievement have been answered, others cannot be answered with available data, or can be answered only by relying on data with serious shortcomings. For example, representative data about the performance of high-achieving college-bound students are meager, leading many analysts to rely instead on unrepresentative, and in many respects misleading, data from college admissions tests. Data about differences in trends among regions are limited and inconsistent, and information about the achievement of students in private schools is extremely scarce.

Improved data from educational achievement tests would therefore clearly be helpful, particularly if test scores continue to serve as a primary rationale for changes in educational policy. If additional data are to be created, the federal government might take responsibility for providing them, and some prominent recent proposals--for example, the Alexander-James report recently published by the Department of Education--have called for expanding federal activities in this area. 1/ The federal role in providing educational statistics is long-standing and largely noncontroversial, and few other organizations have the ability to create data that are nationally representative and consistent over time. On the other hand, tight fiscal constraints would make any increase in outlays difficult.

The findings of this study, however, make a strong case against creating a single "national achievement test" for this purpose. They show clearly that a variety of measures are often needed to reach reasonably certain conclusions about student achievement. Only by comparing several tests can the analyst distinguish results that are consistent enough to provide a firm basis for policy from those that are merely idiosyncrasies of individual tests.

The results of this analysis thus challenge a widespread confidence in the adequacy of individual tests as indices of achievement. Many analysts have relied on one or a few tests--most often, the National Assessment of Educational Progress (NAEP) or the Scholastic Aptitude Test (SAT)--to gauge achievement. Certain proposals to improve data on student achievement--for example, the Alexander-James report, which proposed a major expansion of the NAEP--could further increase the tendency to rely on a single test. Some recent proposals would even eliminate other, independent sources of data by combining them with the NAEP.

The risk of being misinformed by the results of a single test is appreciable, and it is often impossible to foresee when a single test will be misleading. There are certainly many cases in which numerous tests point to similar conclusions, and in such instances a single, high-quality test would be sufficient. But fundamental inconsistencies in results appear relatively often and affect even tests of high quality, such as the National Assessment.

---

1. Lamar Alexander, H. T. James, and others. *The Nation's Report Card: Improving the Assessment of Student Achievement* (Washington, D.C.: Office of Educational Research and Improvement, 1987).

Moreover, the inconsistencies affect even very basic conclusions about trends in achievement--for example, whether trends were more favorable in certain subject areas. Perhaps most important, because some of the significant inconsistencies in the results of current tests were unexpected and remain unexplained, future users of test scores will not always be able to predict when the results of a single test should be accepted with confidence.

Efforts to improve measures of achievement might lessen this problem somewhat but cannot be expected to eliminate it. Inconsistencies in results are probably an inevitable consequence of the incompleteness of any test as a proxy for achievement rather than a sign of remediable flaws in particular tests. Indeed, because some of the important inconsistencies are unexplained, it is not yet clear how tests should be improved to lessen the frequency of these inconsistencies.

Even if it were feasible to eliminate disparities among tests, it would not always be desirable, because those discrepancies can themselves provide important information. Tests often emphasize different skills and knowledge, and disparities in their results can therefore reflect significant differences in students' mastery of various aspects of a subject area.

Appraisals of student achievement thus ideally should be based on a number of diverse measures, both to weed out the misleading, idiosyncratic results of individual tests and to capitalize on meaningful variations in results. This approach, however, imposes difficult trade-offs. The costs of maintaining and improving several tests, for example, would probably limit the improvements made to any one test. Precisely what the compromise should be is open to debate, but some current proposals lean further in the direction of relying on a single test than available data justify.

For certain purposes, it would be important, though costly, to collect data on relevant educational and noneducational factors along with the data from additional educational tests. As noted earlier, the meaning of changes in test scores can depend on the factors that caused them. Collecting data on factors such as dropout rates and demographic changes therefore can be critical.

EVALUATING EDUCATIONAL POLICIES

Much of the current interest in aggregate test scores stems from a desire to determine the success or failure of educational policies. Although aggregate test score data can be useful in this respect, the link between educational policies and aggregate test scores is often far weaker and less straightforward than many observers believe. Data that are adequate as a measure of students' achievement do not always provide a sound basis for evaluating policies.

Simple aggregate trends in test scores, taken alone, are an insufficient basis for evaluating new educational policies. Other factors can markedly influence achievement trends, sometimes more substantially than the specific educational policies at issue, and could even obscure their effects entirely. Beneficial policies, for example, could even be accompanied by falling average test scores. Thus, to appraise new initiatives with confidence, one needs to know how trends are being deflected from the course they would have followed in the absence of those policies, not merely whether scores are rising or falling.

In many instances, assessment of new initiatives will also require data that link test scores to the specific educational experiences of different students. Such data are sometimes needed to eliminate the confusion caused by other, irrelevant influences on test scores. For example, to show that increased course requirements improved achievement, one would want data that indicated particularly favorable trends among students whose course load was altered as a result; positive trends among students whose course load already far exceeded the new requirements would presumably reflect something else. In addition, data linking scores to specific educational experiences are needed to identify differences in the responses of various groups, such as high- and low-achieving students, to a given change in policy.

If simple trends in aggregate test scores are used alone to evaluate new policy initiatives in the near future, they will often overestimate the initiatives' effectiveness. Indeed, they could even suggest a positive effect when initiatives are actually ineffective or moderately harmful. One reason is that the ongoing rise in test scores antedates many of the current initiatives and might have continued in their absence. Even if incoming cohorts of students would not have continued to produce increasing scores, average scores in the higher grades might well have continued rising, since the cohorts that will be entering the higher grades over the coming years have already produced gains in the lower grades. In those instances in which scores would have continued rising even in the absence of policy change, the simple continuation of the rise in scores offers no evidence that an initiative has been effective; rather, success would be indicated only if the rise in scores were augmented.

The effectiveness of programs initiated by states or localities could also be substantially overestimated if the general, nationwide rise in scores is not distinguished from the impact of those specific programs. This error is particularly likely when average scores in a given jurisdiction are compared with national norms that are only infrequently revised (as is the case with most commercial standardized achievement tests). In those instances, when scores are rising nationwide, the typical district or state will see its scores rise relative to the national average simply because the national standard is increasingly out of date and thus progressively lower than it should be.

The effectiveness of some initiatives could also be overestimated because of the tendency by some teachers to "teach to the test"--that is, to tailor their instruction to meet the demands of tests. Both proponents and opponents of the current wave of increased testing agree that greater teaching to the test will result from it. 2/ Regardless of whether this response benefits or harms instruction, it can seriously distort trends in average scores when the instructional goals are much broader than the material tested, which is often the case. In such situations, students' overall achievement can only be gauged fully by using additional measures that capture aspects of the curriculum that are not stressed in the test toward which teachers are directing their instruction.

In other instances, however, simple aggregate trends in test scores will bias evaluations downward, thereby understating or even obscuring the impact of successful educational initiatives. This can happen, for example, in areas where demographic changes in the school-age population are especially rapid. The share of the school-age population comprising historically low-achieving groups--certain minority groups and students with limited (or no) proficiency in English--is rising, as a result of both immigration and differences in fertility among ethnic groups. While these trends are gradual in the nation as a whole, they are much more pronounced in certain jurisdictions, and scores in these areas are likely to be deflected downward from whatever course they would have followed in the absence of these demographic changes.

---

2. See, for example, W. James Popham, Keith L. Cruse, Stuart C. Rankin, Paul D. Sandifer, and Paul L. Williams, "Measurement-Driven Instruction: It's on the Road," *Phi Delta Kappan*, vol. 66 (May 1985), pp. 628-634; and R. M. Jaeger, "The Final Hurdle: Minimum Competency Achievement Testing," in G. R. Austin and H. Garber, eds., *The Rise and Fall of National Test Scores* (New York: Academic Press, 1982), pp. 223-246.

Simple trends in test scores will also underestimate the success of new policies if those initiatives are accompanied by certain changes in the selection of students for testing. As explained earlier, selection changes can substantially--and deceptively--alter average test scores, and some educational initiatives could depress scores by altering selection even while improving achievement. The most obvious instance would be initiatives that lowered the dropout rate. Because students who drop out score on average below others, their retention in school could depress average scores or attenuate an ongoing rise, even if their own scores rose as a result of remaining in school. Ironically, the negative effect on test scores--and the resulting underestimate of the program's effectiveness in raising achievement--would be proportional to the programs' success in lowering the dropout rate. Similar distortions could also arise in other ways--for example, if a new program reduced the frequency of unnecessary assignments to special education programs and thereby retained additional low-scoring students in the group routinely tested. 3/

IMPROVING EDUCATIONAL ACHIEVEMENT

Over the last decade, trends in test scores and views about their causes have provided a basis for formulating new educational policies and for presuming their effectiveness. Many people have assumed that a few key factors responsible for much of the decline of the 1960s and 1970s could be identified, and that simply reversing those variables would bring about a similarly dramatic improvement in scores.

This study, however, offers scant encouragement to those who would search among the causes of the recent trends for a few key factors that might cause major improvements in achievement. Although educational factors of that potency might exist, the analysis of past trends reported here did not identify them. On the contrary, if the evidence about the recent past is to serve as a guide, it suggests that modest expectations about the impact of individual educational changes are appropriate. The individual effects of the educational factors that contributed to the achievement trends of the past two decades were small compared with the total change in average scores. Indeed, the substantial contribution of noneducational causes to the recent trends indicates that the total effect of all educational causes combined--including those not assessed in this

---

3. By the same token, the apparent effectiveness of policies could be exaggerated by manipulating selection to exclude lower-scoring students from the group routinely tested.

study--fell considerably short of the total change in scores. Thus, to bring about comparably large and pervasive improvements in scores in the future would require a significantly more potent mix of educational changes--including a greater number of factors, more powerful factors, or more drastic changes in certain factors--than was involved in the trends of the past two decades.

The results of this study therefore suggest searching broadly for factors that may improve achievement. Restricting new initiatives to factors that can be linked to past trends could be counterproductive, not only because the impact of those factors would often be smaller than hoped, but also because other factors with equal or greater potential might be ignored. For example, among the factors whose contributions to past trends cannot be gauged because data are inadequate are some--such as students' attitudes and motivation, demands for writing in the classroom, and local graduation requirements--that might have a major impact on students' learning. Even certain of the factors that apparently did not contribute to recent trends--specifically, those that are temporally inconsistent with the trends but that can affect achievement more generally--might nonetheless prove important in the future. For example, the finding that state graduation standards apparently did not contribute to the latter half of the achievement decline does not imply that increases in those requirements will prove ineffective later. The finding that such factors did not contribute to the recent achievement trends merely removes one basis for presuming their effectiveness.

Indeed, the results of this analysis suggest that the effectiveness of the current wave of initiatives should not be presumed on the basis of assumptions about what caused past trends. In many ways, the initiatives are more appropriately seen as an experiment than as a clear-cut response to the trends of the past two decades, and discerning the effects of the initiatives--both beneficial and detrimental--will require careful evaluation.

Even though analysis of past trends does not point to the few key factors that many analysts have wanted to find, it can be useful in focusing new initiatives. For example, the abundant instances in which many students are failing to master knowledge and skills that most people would consider fundamental provide ample suggestions of areas in which instruction needs strengthening. These weaknesses are apparent in diverse subject areas, ranging from knowledge about American government to the ability to apply fundamental mathematics to problems of everyday life.

To successfully counter some of the most troubling aspects of recent data on test scores, initiatives would have to focus on higher-order skills. The term "higher-order" can be used in various ways, but here it refers to skills--such as inferential comprehension in reading, problem-solving, and other applications in mathematics--that entail substantial reasoning and cannot be learned by rote.

While higher-order skills are clearly a more significant aspect of achievement in the higher grades and in the particularly complex material that the highest-achieving students are expected to master, they are also important even in the case of some rudimentary material. Many skills that are "basic"--in the sense of being simple, fundamental skills that all students are expected to master--are nonetheless "higher order" in that they entail reasoning, problem-solving, and so on. Examples include the ability to solve simple word problems involving percentages and the application of arithmetic algorithms to such problems as comprehending utility bills. Proficiency in writing, which many would consider a basic skill--it is one of the "three Rs"--might also fall into this category, for it too involves cognitive skills more complex than the rote learning of facts and algorithms.

Indeed, despite the particularly serious problems in higher-order skills and the greater decline in the higher grades, initiatives that ignore the lower grades--and some have--would miss some of the most important problems revealed by the achievement data. Many of the most troubling deficiencies, including those involving higher-order skills, appear in material taught in the elementary and junior high grades. If the smaller upturn to date in the higher grades is misunderstood as being a fundamentally slower rate of improvement in those grades, it might be seen as a reason to shift emphasis further toward the secondary level despite the existence of these problems in the elementary grades. The smaller rise in scores in the higher grades, however, now appears to be largely an artifact of the smaller number of improving cohorts that have reached the higher grades and not a sign of less rapid improvement.

Ideally, then, educational changes must tread a thin line, strengthening rudimentary skills in many areas without allowing an overemphasis on basic skills that would crowd out instruction in higher-order skills. While striking this balance would be important in any case, the serious erosion of higher-order skills in the recent past make it all the more so. Precisely where that line lies is a matter of judgment, but many observers feel that certain curriculum changes during the past decade and a half have overemphasized "basics." An expert panel convened to assess the implications of the National Assessments of mathematics, for example, argued that a back-to-basics orientation--specifically, emphasis on computation, facts, and definitions at the expense of problem-solving--narrowed the mathematics curriculum and thus contributed to the particularly severe declines observed in higher-order skills in mathematics in the 1970s. 4/ A current tendency to refer to even higher-level curriculum requirements as "basic"--for example, the labeling by the National Commission on Educational Excellence of high school mathematics (including algebra, geometry, and elementary statistics) as one of the "New Basics"--could inadvertently cloud this critical issue.

Recent data also suggest the importance of focusing on the education of certain traditionally lower-scoring groups, both because their average achievement remains disturbingly low and because of the promising gains some groups, such as black and Hispanic students, have recently made. Simply assuming that educational initiatives directed toward the student body as a whole will have the intended effects with low-achieving students as well risks eroding their recent gains. These gains remain largely unexplained, because the commonly cited explanations--desegregation and federally funded compensatory education--can account for only a moderate share of the improvement. Until some of the other factors that helped bring about these gains have been identified, there is a substantial risk that policies contributing to the gains might be inadvertently weakened or abandoned as a side effect of more general efforts to improve education. Careful monitoring of the effects of policy initiatives on the achievement of these specific groups of students are needed, and new policies might require alteration, if these gains are to be augmented.

---

4. National Assessment of Educational Progress, *Trends in Mathematical Achievement, 1973-78* (Denver: NAEP/Education Commission of the States, August 1979), p. 25.

APPENDIX

THE EFFECTS OF SPECIFIC FACTORS ON ACHIEVEMENT TRENDS

This appendix summarizes the evidence pertaining to the contributions of over two dozen specific factors to test score trends. Societal, educational, and selection factors are discussed in separate sections.

SOCIETAL FACTORS

The societal factors considered here are extremely diverse, since this category is a residual that includes any factors that are neither educational nor selection variables. The category includes changes in the ethnic composition of the school-age population, various trends in household and family composition, students' attitudes and behavior, and environmental factors.

Changes in the Ethnic Composition of the Entire Cohort

The percentage of minority students in the school-age cohort has been growing, and since the groups accounting for much of that growth have, on average, substantially lower achievement scores than do nonminority students, this shift contributed to the achievement decline and impeded the subsequent upturn. These changes in ethnic composition have been gradual and slight, however, and their effects on recent achievement trends have been correspondingly small. 1/

---

1. The term "ethnicity" as used here encompasses some distinctions--such as that between blacks and whites--that are often popularly termed racial. The ethnic categories used here are based on, but differ substantially from, those used by the Bureau of the Census. Specifically:

   - "Black" refers to all individuals who are so identified in the Current Population Survey by the respondent in the household, except for those who also identify themselves as Hispanic.
   - "Hispanic" refers to all individuals who are identified as being of Hispanic origin or descent, regardless of race. The vast majority of Hispanics are also identified as white.
   - "Nonminority" refers to those who are neither black nor Hispanic, as defined above, and who do not identify themselves as members of other minorities (such as Native Americans or Asians).

Between the 1971 and 1979 school years (from the first year in which comparable data were available until the approximate end of the decline in achievement among high school seniors), the minority share of the total school-age population (ages 6 through 17) increased from 21 percent to 25 percent. 2/ Hispanic students accounted for roughly half of the total increase in the minority share, black students for about 30 percent, and other minority students (mostly Asians) for the remaining fifth. (The percentages among senior-high students were similar, although the minority proportion was a bit lower in that age group.)

The impact of this shift would vary from one test to another, because disparities in scores among ethnic groups differ markedly among subject areas and tests. For example, Asian students taking the SAT score above the nonminority average on the mathematics scale but below the nonminority average on the verbal scale. 3/ Moreover, the disparity between minority groups and nonminority students can change as the minority groups grow--especially when immigration is a major source of the increase, for the new members of the group can differ substantially from previous cohorts. Current Asian immigrants, for example, represent a different mix of ethnicities than did the Asian students of the recent past and should not be expected to show the same achievement patterns. 4/

In the nation as a whole, changes in the ethnic composition of the school-age population between 1971 and 1979 probably depressed the score of the median student by roughly one percentile--that is, from the 50th to the 49th percentile--or even less, depending on the test. By comparison, during the same period, drops of five to nine percentiles were observed on some tests administered to high school seniors, and the SAT-Verbal dropped by 11 percentiles.

Single-Parent Households

The proportion of children living in single-parent households has grown markedly over the past 25 years. Whatever the effect on the achievement

---

2. Congressional Budget Office tabulations of the March Current Population Survey, 1972 through 1980.

3. College Entrance Examination Board, "College Board Data Show Class of '85 Doing Better on SAT, Other Measures of Educational Achievement" (New York: The College Board, press release, 1985).

4. See Robert W. Gardner, Bryant Robey, and Peter C. Smith, "Asian Americans: Growth, Changes, and Diversity," *Population Bulletin*, vol. 40, no. 4 (October 1985), Tables 1 and 9.

of individual children, however, this trend apparently contributed only trivially to the overall decline in test scores.

A number of cross-sectional studies have found that children from single-parent households headed by women have lower average scores on a number of measures of intellectual development and achievement, including IQ tests, standardized achievement tests, and school grades. [5] This association between number of parents present and achievement varies markedly from one study to another, however, depending in part on the characteristics of the children involved. For example, a recent nationally representative study found that the scores of elementary school children from two-parent households exceeded those of children from single-parent homes by roughly 0.13 standard deviation among whites and 0.20 standard deviation among blacks. [6] In contrast, the corresponding differences among secondary school students were found to be negligible in a parallel study of that age group. [7]

The impact of the growing share of children living in single-parent households might be even less than these cross-sectional findings suggest, however, because the general problem of confounding is particularly acute in this instance. For example, school-age children in female-headed households are more than four times as likely as other children to be

---

5. See, for example, E. M. Hetherington, K. A. Camara, and D. A. Featherman, "Achievement and Intellectual Functioning of Children in One-Parent Households," in J. T. Spence, ed., *Achievement and Achievement Motives: Psychological and Sociological Approaches* (San Francisco: W. H. Freeman, 1984); A. M. Milne, D. E. Myers, F. M. Ellman, and A. Ginsburg, "Single Parents, Working Mothers and the Educational Achievement of Elementary School Age Children" (Washington, D.C.: Decision Resources, unpublished, June 1983); D. E. Myers, A. Milne, F. Ellman, and A. Ginsburg, "Single Parents, Working Mothers and the Educational Achievement of Secondary School Age Children" (Washington, D.C.: Decision Resources, unpublished, June 1983); Sally Banks Zakariya, "Another Look at Children of Divorce: Summary Report of the Study of School Needs for One-Parent Families," *Principal* (September 1982), pp. 34-37; and D. Scott-Jones, "Family Influences on Cognitive Development and School Achievement," in E. W. Gordon, ed., *Review of Research in Education,* vol. 11 (Washington, D.C.: American Educational Research Association, 1984), pp. 259-304.

6. Milne and others, "Single Parents, Working Mothers and the Educational Achievement of Elementary School Age Children."

7. Myers and others, "Single Parents, Working Mothers and the Educational Achievement of Secondary School Age Children"; Zakariya, "Another Look at Children of Divorce."

poor. 8/ In addition, minority school-age children are more than two-and-one-half times as likely as nonminority children to live in female-headed households. 9/ Much of the research showing lower achievement among children from female-headed households fails to control adequately for such factors, and several studies that have taken those factors into account have found that the apparent differences between children from female-headed and other households shrink as a result. 10/ How much of the apparent achievement gap between children from single-parent and two-parent families to attribute to that aspect of family composition itself remains a matter of controversy, and therefore how much one should expect trends in the percentage of children living in single-parent families to affect average test scores is correspondingly uncertain. 11/

Since the 1959-1960 school year, the proportion of children living in single-parent, female-headed households has grown from 9 percent to about 20 percent. 12/ (The proportion of children living in single-parent, male-headed households has also grown, but that percentage remains small--about 2 percent in 1984.) 13/ Although this trend was virtually uninterrupted until the last few years, the most rapid increase occurred between 1969

---

8. Congressional Budget Office, "Poverty Among Children" (Staff working paper, December 3, 1984).

9. Based on Bureau of the Census, *Current Population Reports*, Series P-60.

10. Hetherington and others, "Achievement and Intellectual Functioning"; D. Scott-Jones, "Family Influences on Cognitive Development and School Achievement."

11. On the other hand, the cross-sectional studies quite likely understate--perhaps by a large margin--the impact of living in a single-parent household on the achievement of certain individual children. It is reasonable to assume that any effect on achievement increases with the time that children live in single-parent households, but cross-sectional data typically include little or no indication of that duration and therefore probably obscure the greater effects on children living in single-parent households for long periods. While such an understatement would be important in some contexts, it is not germane here, for the national trend data on household composition parallel the cross-sectional data in grouping children together regardless of the duration of their time in single-parent homes.

12. These percentages are from Bureau of the Census, *Current Population Reports*, Series P-60; they include only related children in families. Trends among school-age children have been similar in recent years, although the proportion in female-headed households is somewhat higher.

13. Department of Commerce, Bureau of the Census, *Marital Status and Living Arrangements, 1984*, Series P-20, No. 399 (1985), Table 4.

and 1977 and was thus roughly concurrent with the decline in achievement.[14] Taken alone, this timing suggests that the growing share of children living in single-parent households could have contributed to the achievement decline.

Because of the relatively small number of children directly affected by the trend, however, the growing proportion of children living in single-parent households could have contributed only trivially to the test score decline. The great majority of children remain in two-parent households, and their scores would not be directly affected. For example, between 1965 and 1979, the proportion of school-age children living in female-headed households increased by only about eight percentage points, leaving the scores of 92 percent of the students in those cohorts unaffected. If the effect of being in a single-parent home was to depress the test scores of affected children by an average of 0.15 standard deviation, this shift in household composition would have lowered the overall average test score by roughly 0.01 standard deviation. In contrast, declines in average scores in excess of a third of a standard deviation were not uncommon during that period. Moreover, in secondary schools--where the test score decline was typically largest--the contribution of this shift in household composition would be smaller yet or even nonexistent.

Family Size

The fertility changes of the baby boom and subsequent baby bust produced several changes in the composition of families that can be conveniently--if not entirely accurately--grouped together as changes in "family size." The baby boom raised the average number of children per family and the average birth order of children. [15] The baby bust reversed both of these trends.

---

[14] The extent of temporal consistency is not fully apparent, because this trend cannot be linked precisely to birth cohorts and cannot be aligned with test score trends in specific grades. The trend data also provide no indication of the pattern of household composition experienced over time by affected children--for example, the ages at which children encounter various household arrangements--which is an important omission, since factors such as age appear to alter markedly the effects on achievement.

[15] "Children per family" is used here to denote the average number of resident children under age 18 per family; families with no resident children under age 18 are not averaged in. Birth order refers to the sequence of births in a family; the first-born has an order of one, the second-born, two, and so on. "Average birth order" in a cohort is simply the average order of all children born in that year. If half are first-borns and half second-borns, their average birth order is 1.5; if a third each are first-, second- and third-borns, their average birth order is 2.0, and so on.

A well-publicized but still controversial hypothesis attributes a sizable share of both the decline in test scores and the subsequent rise to these changes. Indeed, one researcher used trends in birth order to predict quite accurately the time when SAT scores would start their upturn and has since offered predictions of trends in SAT scores past the year 2000. 16/

The prominence of this hypothesis probably stems less from the long-standing and copious research into the effects of family size on intelligence and achievement than from the striking concordance between trends in average birth order and SAT scores over the past two decades (see Figure A-1). Average birth order rose steadily from 2.4 to 3.0 between the cohort born in 1947 and those born in 1961 and 1962--almost exactly the cohorts that produced the decline in SAT scores. Both trends have since reversed themselves--birth order sharply, SAT scores more modestly.

The research as a whole suggests that family size could have contributed to both the decline and the rise of test scores. Despite the striking consistency between trends in birth order and SAT scores, however, changes in family size appear to account for only a modest share of the trends in test scores.

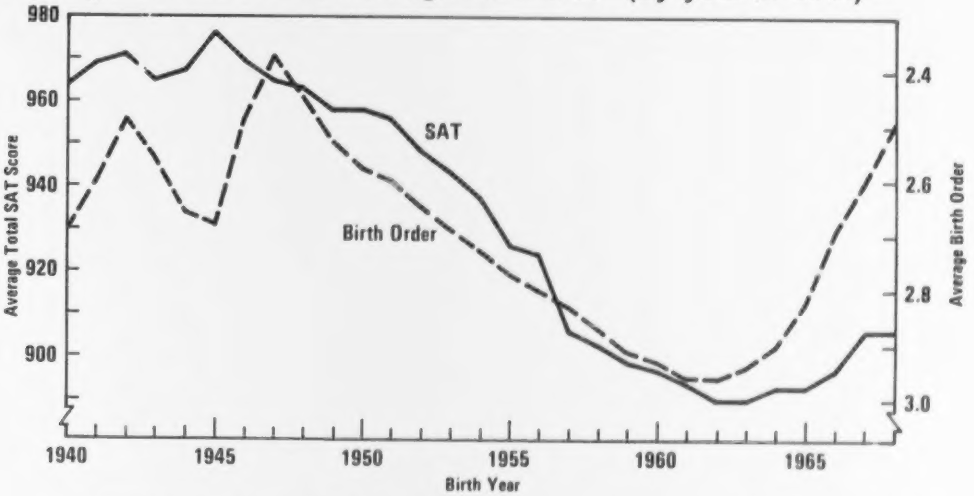
This conclusion, however, does not represent a consensus in the research literature. Indeed, research on this topic is currently characterized by vehement disagreements, and the available cross-sectional research and data on temporal consistency are used to support a wide range of contradictory positions. For these reasons, this analysis gives special weight to a few studies that directly estimated the contributions of family size to recent achievement trends by comparing the family characteristics and test scores of individual students in some of the cohorts responsible for those trends. Those studies are described after the following synopsis of cross-sectional studies and temporal consistency.

---

16. R. B. Zajonc, "Family Configuration and Intelligence," *Science*, vol. 192 (April 16, 1976), pp. 227-236, and "The Decline and Rise of Scholastic Aptitude Scores: A Prediction Derived from the Confluence Model," *American Psychologist*, vol. 41, no. 8 (August 1986), pp. 862-863. Whatever the impact of birth order on achievement, such a prediction assumes that all other factors affecting aggregate achievement will vary little over the coming years.

Figure A-1.

Average Total SAT and Average Birth Order (By year of birth)



SOURCES: Congressional Budget Office calculations based on Hunter M. Breland, *The SAT Score Decline: A Summary of Related Research* (New York: The College Board, 1987); The College Entrance Examination Board, *National College-Bound Seniors, 1985* (New York: The College Board, 1985); and National Center for Health Statistics, unpublished data.

NOTE: Birth order is inverted so that trends in birth order and SAT scores are in the same direction.

Cross-Sectional Studies. The relationships between family or household composition and various aspects of intelligence and achievement have been noted for at least a century, although the nature of those relationships and their explanations remain controversial to this day. 17/ The association between achievement and the number of children has probably received the greatest attention, but studies of birth order are also abundant, and some prominent analysts have treated the two variables--incorrectly, as is

---

17. For example, Francis Galton, *English Men of Science* (London: MacMillan, 1874) cited in Joseph Lee Rodgers, "Confluence Effects: Not Here, Not Now," *Developmental Psychology*, vol. 20 (1984), pp. 321-331.

explained below--as roughly synonymous. Other related changes in family composition have received less attention and are not considered here. 18/

Available research leaves no doubt that the factors termed "family size" in this study, taken together, are associated with achievement in most settings. What remains controversial is which aspects of family size are important and what causes these associations. Without answers to these questions, the contribution of changing family size to recent trends in test scores cannot be accurately assessed.

Most cross-sectional research shows that children from larger families tend on average to leave school earlier and to score lower on intelligence and achievement tests than their peers from smaller families. 19/ This relationship has been found in many different groups in several countries, and it seems to hold true for a wide variety of measures of intelligence, educational achievement, and educational attainment.

The relationship between birth order and achievement is less certain. Studies that attempt to isolate an independent effect of birth order--typically, by examining the relationship between birth order and achievement among families with a specific number of children--are inconsistent. Some studies show an independent negative association between birth order and achievement, while others do not. Some analysts suggest that this inconsistency reflects different effects in different age groups: among older children, later-born children generally score lower than earlier-born, while the pattern among younger children is less clear and may even be re-

---

18. One other aspect of family composition that warrants special note is the spacing between births. It has been argued that the effects of birth order and family size are mediated by changes in this factor (R. B. Zajonc, "Validating the Confluence Model," *Psychological Bulletin*, vol. 93 (1983), pp. 457-480). However, research directly assessing the impact of spacing on achievement or IQ (rather than attempting to infer it from data on trends in other family characteristics) suggests that while spacing affects performance, it does not substantially alter the relationship between number of children and performance (see Yvonne Brackbill and Paul L. Nichols, "A Test of the Confluence Model of Intellectual Development," *Developmental Psychology*, vol. 18 (1982), pp. 192-198). Therefore, omitting spacing of births from this discussion should not bias conclusions about the effects of birth order and number of children.

19. For reviews of many of the relevant cross-sectional studies, see R. B. Zajonc, "Validating the Confluence Model"; Judith Blake, "Family Size and the Quality of Children," *Demography*, vol. 18 (November 1981), pp. 421-442; Rodgers, "Confluence Effects: Not Here, Not Now"; and Lala Carr Steelman, "A Tale of Two Variables: A Review of the Intellectual Consequences of Sibship Size and Birth Order," *Review of Educational Research*, vol. 55, no. 3 (Fall 1985), pp. 353-386.

versed. 20/ Not all studies of older children, however, have shown a consistent independent association between birth order and achievement.21/

Researchers have reached fundamentally different conclusions about the causes of these associations between family size and achievement. The primary root of the disagreement is a particularly serious instance of the common problem of confounding. Family size is usually related to other factors, such as ethnicity and socioeconomic status (SES), that are in turn strongly associated with educational achievement. In the United States as a whole, for example, the average number of children per family was roughly 1.8 among whites, 2.2 among Hispanics, and 1.9 among blacks in 1984.22/ Similarly, families with a greater number of children are headed by parents who have on average lower educational attainment and lower occupational prestige. 23/

The extent to which the associations found in cross-sectional studies should be attributed to these confounded factors rather than to family size itself remains a matter of intense controversy. Some researchers argue that apparent effects of family size are primarily consequences of associated

---

20. R. B. Zajonc, Hazel Markus, and Gregory B. Markus, "The Birth Order Puzzle," *Journal of Personality and Social Psychology*, vol. 37 (1979), pp. 1325-1341; and Zajonc, "Validating the Confluence Model," Figures 2, 3, 5, and 6.

21. In nationally representative data from the high school senior class of 1972, for example, negative associations between achievement and birth order appear in families with two or three children, but not in those with four or five. See Albert E. Beaton, Thomas L. Hilton, and William B. Shrader, *Changes in the Verbal Abilities of High School Seniors, College Entrants, and SAT Candidates Between 1960 and 1972* (New York: College Entrance Examination Board, June 1977), Table 10. See also Steelman, "A Tale of Two Variables."

22. Department of Commerce, Bureau of the Census, *Household and Family Characteristics: March 1984*, Current Population Reports: Population Characteristics, Series P-20, No. 398 (1985), Table 1. Hispanics are counted twice in these numbers because the Census Bureau asks about race independently of questions on ethnic origin. The average number of children in the "white" category would drop if Hispanics were excluded.

23. Blake, "Family Size and the Quality of Children"; Judith Blake, "A Sociological Perspective on Number of Siblings and Educational Attainment" (paper delivered at the annual meeting of the American Association for the Advancement of Science, May 27, 1985); Brackbill and Nichols, "A Test of the Confluence Model"; Ellis B. Page and Gary M. Grandon, "Family Configuration and Mental Ability: Two Theories Contrasted with U. S. Data," *American Educational Research Journal*, vol. 16, no. 3 (Summer 1979), pp. 257-272; Rodgers, "Confluence Effects: Not Here, Not Now"; and Zajonc, "Validating the Confluence Model."

differences in factors such as ethnicity. 24/ Researchers at the other extreme argue that some sizable proportion of the observed relationships are in fact direct effects of family characteristics. 25/

The likely contribution of changes in family size to the achievement trends of the past two decades hinges largely on the extent to which each of these competing views is correct. To whatever degree family size itself caused the observed cross-sectional relationships, the changes in family size accompanying the baby boom and baby bust should have brought about corresponding changes in achievement, regardless of confounding with variables such as socioeconomic status and ethnicity. If, on the other hand, the confounded variables account for some or all of the observed relationships, the effects of the baby boom would have been that much smaller, because the fertility changes of the baby boom did not cause parents to change in terms of factors such as educational attainment and ethnicity.

Perhaps the only noncontroversial conclusion that can be drawn from this research is that confounded factors account for an appreciable share of the observed relationships between family size and achievement. This conclusion implies that the cross-sectional research overstates the likely contribution of changes in family size to recent achievement trends, but the magnitude of that overstatement remains unresolved.

Temporal Consistency. In certain respects, trends in family size show a remarkable consistency with some aspects of the achievement trends, but in other respects, they are inconsistent. Taken together, the data about temporal consistency certainly do not rule out family size as a contributor to the achievement trends, but they are not nearly as striking or persuasive as some observers have maintained.

As noted above, trends in average birth order show a striking consistency with achievement trends during the later years of the decline and the subsequent upturn. This consistency is not limited to the SAT. Among a variety of tests, the end of the achievement decline and the onset of the subsequent rise in test scores occurred within a few years of the birth cohorts of 1962 and 1963--that is, very nearly at the point at which birth order began falling. On the other hand, trends in birth order are far less temporally consistent with the early years of the achievement decline. The

---

24. Page and Grandon, "Family Configuration."

25. For example, Blake, "Family Size and the Quality of Children"; and Zajonc, "Validating the Confluence Model."

beginning of the decline did not show a cohort pattern at all, and the birth cohorts that initiated the decline ranged from 1946 (a few cohorts before birth order began to rise) to 1956 (by which time the rise in birth order was nearly over). One possible explanation for this pattern is that trends in birth order contributed to the achievement trends, but that their influence was modest enough to be offset by other factors during the early years of declining scores.

The cross-sectional research, however, suggests that birth order is less important than the number of children per family. Given the constraints of available data, the temporal consistency between number of children per family and test scores is hard to gauge, but it appears not to be as close as that shown by birth order.

At first glance, the trend in the average number of children per family appears to be entirely inconsistent with test score trends. The average number of children rose from about 2.2 in 1953 (the earliest year of data) to about 2.4 in 1965 as a consequence of the baby boom. It has fallen quite consistently since then, although the drop has tapered off recently. By 1984, the average number of children per family was only 1.85. 26/ Thus, the drop--which should have raised test scores--continued almost without interruption during the entire period of the test score decline and began to abate only recently, at a time when test scores were generally rising.

In fact, however, trends in the number of children per family are not as inconsistent with achievement trends as they first seem. The trend data about the number of children per family discussed in this analysis were obtained by surveys that inquired about all children under age 18 living in the household at the time of the survey (in March of every year). Each year's average thus reflects children of 18 different ages--that is, 18 different birth cohorts, ranging from the cohort born in the year of the survey to that born 17 years earlier. When the average number of children per family reached its peak in 1965, for example, that year's data reflected cohorts born from 1948 to 1965.

Data on the average number of children per family in any given year therefore cannot be tied to individual birth cohorts. As long as the average number of children per family is changing, each of the cohorts reflected in a given year's data will experience a different history of family sizes over the course of their childhoods. For example, the birth cohort of 1953 exper-

---

26. Department of Commerce, Bureau of the Census, published and unpublished tabulations of the March Current Population Survey.

ienced increasing family sizes--from 2.2 to 2.4 children, on average--during the first 12 years of life. From then until age 18, they experienced the rapid decline in family size--from 2.4 to 1.9 children, on average--that appears in the survey data beginning in 1965. In contrast, the birth cohort of 1965 experienced that same decline in family size during the early years of childhood.

Because the simple trend in the average number of children reflects children of all ages who experienced the change for varying lengths of time and at different periods in their childhood, it does not provide the information needed to gauge the contribution of family size to trends in test scores. Instead, one needs data indicating the number of siblings present in the home of children throughout their childhoods, as well as a model indicating which periods of childhood are most susceptible to the influence of family size. These data do not exist. Moreover, they cannot be derived easily from the available information about family size at specific points in childhood, such as birth order (which is closely related to the number of siblings present at a child's birth) or the number of siblings present at the conclusion of schooling. 27/

If the ideal data were available, however, they would probably be more consistent with trends in test scores than is the trend in average number of children. The ideal measure for high school seniors, for example, would probably predict a gradually growing, positive effect of family size on test scores that became substantial either in the later years of declining test scores or during the period of rising scores. To understand this, consider the experience of successive cohorts of 17-year-old students as the average number of children per family fell. When the decline in fertility caused the average number of children to begin falling in the mid-1960s, the cohort that was then age 17 would have been little affected. The number of 17-year-olds with newborn siblings would be changed very little, and for the few whose circumstances were altered, the change would be confined to the last year of childhood. With each passing year, the number of 17-year-olds influenced by the change would grow, and the portion of their childhood affected would increase.

Direct Estimates of the Impact of Changing Family Size. Studies based on data about individual students in affected cohorts suggest that trends in birth order and number of children per family produced only a small to moderate share of the test score decline. Moreover, these studies, like the

---

27. For example, for the next-to-last-born of 10 children, the average number of minor children present at age 17 is two--hardly an accurate indication of the family configuration experienced by that child during most of his or her childhood.

cross-sectional studies mentioned earlier, overstate the independent effects of changes in family size, because they failed to take into account the effects of confounded variables such as socioeconomic status and ethnicity.[28] No studies to date have used individual-level data to estimate the contribution of these trends to the subsequent rise in scores.

One study that examined changes in birth order in the entire age cohort estimated that between 1964 and 1976, changes in birth order would have produced a drop of 6.3 points on the SAT-Verbal--about 15 percent of the total observed decline. [29] A shorter-term but more detailed study examined changes in both number of children and birth order among students actually taking the SAT. That study estimated that between 1970 and 1976, roughly 4 percent of the decline on the SAT-Verbal and 9 percent of the drop on the SAT-Math could be attributed to changes in these factors.[30] The proportion of the decline attributable to these factors might have been greater, however, among all students than it was among those taking the SAT. One study found that between 1959 and 1971, about 25 percent of the decline in reading achievement among all seniors could be attributed to changes in number of children and birth order, compared with 9.5 percent of the decline on the SAT-Verbal. [31]

Conclusion. Given the inconsistencies in the research discussed above, it is perhaps not surprising that analysts have reached sharply different conclusions about the contributions of family size to the recent trends in test scores. If one focuses on birth order, one finds clear temporal consistency with trends in test scores for about two decades (though inconsistency in earlier years), but ambiguous cross-sectional research. The evidence pertaining to the average number of children per family is the reverse: clear cross-sectional evidence, but amibiguous evidence about temporal consistency. Uncertainty about the causes of the cross-sectional relationships further increases the ambiguity of the data as a whole.

---

[28] One of these studies attempted to remove the effects of one aspect of socioeconomic status - - family income (R. B. Zajonc and J. Bargh, "Birth Order, Family Size, and Decline of SAT Scores," *American Psychologist*, vol. 35 (July 1980), pp. 662-668). Removing that one factor, however, is insufficient to ascertain how much the estimated effects of family size--small in any case in that study--would have been lessened if ethnicity and more varied indicators of socioeconomic status had also been examined.

[29] H. M. Breland, *Family Configuration and the Decline in College Admissions Test Scores: A Review of the Zajonc Hypothesis* (New York: College Entrance Examination Board, January 1977).

[30] Zajonc and Bargh, "Birth Order, Family Size, and Decline of SAT Scores."

[31] Beaton and others, *Changes in the Verbal Abilities of High School Seniors*, pp. 5, 31, and 57. The proportionately lesser impact of these changes on average SAT scores might reflect the major compositional changes affecting the SAT during those years. Those compositional changes presumably contributed more to the total decline on the SAT than to the amount of the SAT decline attributable to changes in family size.

Faced with these uncertainties, this analysis relies heavily on the few studies that directly estimated the contribution of family size to the recent trends by examining the scores of individual students in affected cohorts. The results of these studies also varied considerably, but they are all consistent with the general conclusion that the contribution of family size was in the range of small to moderate. All of these studies were restricted to portions of the period of declining scores, but it seems plausible that the effects of family size continued during the period of rising scores as well.

Maternal Employment

In 1970, 43 percent of all school-age children (ages 6 through 17) had mothers in the labor force; 15 percent of those children were in families maintained by women. By 1984, the proportion with mothers in the labor force had risen to 60 percent; 23 percent of those children were in families headed by women. 32/

While this trend has often been suggested as a cause of the decline in test scores, the available cross-sectional data do not provide consistent evidence that maternal employment has a negative effect on achievement. A recent review by a National Academy of Sciences panel concluded that "existing research has not demonstrated that mothers' employment per se has consistent direct effects, either positive or negative, on children's development and educational outcomes." 33/ Three other studies using large, nationally representative samples found conflicting results in this regard: two studies found negative effects of maternal employment on achievement in some groups of children, while the third found that children

---

32. Department of Labor, *Perspectives on Working Women: A Databook* (October 1980), Table 30; H. Hayghe, "Working Mothers Reach Record Number in 1984," *Monthly Labor Review*, vol. 107 (December 1984), pp. 31-33. "Families maintained by women" are defined here as those headed by widowed, divorced, and never-married women, as well as married women with absent spouses.

33. C. D. Hayes and S. B. Kamerman, eds., *Children of Working Parents: Experiences and Outcomes* (Washington, D.C.: National Academy Press, 1983), p. 221.

of working mothers in the aggregate showed higher average achievement in reading than other children.[34]

The inconsistent findings of cross-sectional studies is not surprising. There is no reason to expect that maternal employment represents uniform experiences for children or that it would directly affect educational outcomes. Instead, maternal employment is presumably germane to achievement through its effects on other factors--such as the amount and type of interaction children have with parents, the interaction they have with other children, the time spent with other adults in the home or in other diverse settings, the amount of family income, the types of stresses experienced by children, and so on. The simple fact of maternal employment or lack of employment provides little information about these more directly relevant factors.

In addition, maternal employment is seriously confounded with other characteristics of families (such as ethnicity, family composition, age, and income), and existing research is not fully adequate to disentangle its effects. For example, while the newest National Assessment found higher levels of reading achievement among the children of working mothers, it also found that mothers working outside the home had higher average levels of education.[35] Greater parental education is itself associated with higher levels of children's achievement, and this association may underlie the apparent effects of maternal employment.

Some studies have found that the association between maternal employment and achievement also varies depending on socioeconomic status (SES), ethnicity, and the presence or absence of the father. For example, a number of studies have found maternal employment to be associated with higher achievement in families with low socioeconomic status but to be either unrelated or negatively related to achievement in middle-class families.[36] Simple aggregate statistics on national trends in maternal

---

34. Negative effects of maternal employment in some subgroups were found in A.M. Milne, D. E. Myers, F. M. Ellman, and A. Ginsburg, "Single Parents, Working Mothers and the Educational Achievement of Elementary School Age Children" (Washington, D.C.: Decision Resources, unpublished, June 1983); and D. E. Myers, A. Milne, F. Ellman, and A. Ginsburg, "Single Parents, Working Mothers and the Educational Achievement of Secondary School Age Children" (Washington, D.C.: Decision Resources, unpublished, June 1983). Higher average reading achievement among students with working mothers was found in the most recent National Assessment of Reading (National Assessment of Educational Progress, *NAEPGRAM 3: The Home*, Spring 1985).

35. National Assessment of Educational Progress, *NAEPGRAM 3*.

36. Scott-Jones, "Family Influences."

employment, which do not distinguish between these groups, therefore do not indicate what the effect on achievement might be.

The temporal consistency of trends in maternal employment and changes in test scores is hard to gauge, because aggregate data on maternal employment are difficult to align with specific birth cohorts. It is noteworthy, however, that while the growth in maternal employment continued during the entire period of declining test scores, it also persisted after scores began rising again. This need not indicate that trends in maternal employment were unimportant, but it does suggest that if they did contribute to the achievement decline, their effects were small enough to have been overcome by other factors in recent years.

Students' Use of Alcohol and Other Drugs

The widespread changes in students' use of alcohol, marijuana, and other drugs that accompanied the cultural trends of the 1960s and 1970s might have affected achievement in two ways. Changes in drug use might directly affect achievement, and they could also be a marker of other changes--in students' attitudes toward success in school, for example--that also affected performance in school.

The limited data on temporal consistency suggest that changes in drug and alcohol use could have contributed to some aspects of the achievement trends since the early or mid-1970s--perhaps to the greater severity or longer duration of the decline in the higher grades. Heavy use by high school seniors of the most commonly used drugs--tobacco, marijuana/hashish, and alcohol--appears to have crested in the school years 1976 through 1978. The proportion of seniors reporting daily use of marijuana or hashish during the previous month, for example, rose from 6 percent in school year 1974 to 11 percent in 1977 and dropped to 5 percent in 1983. 37/ Earlier data about less frequent use indicate a steady rise in drug use earlier in the 1970s and suggest similar trends among youth as young as 14 and 15.38/

Trends in drug use could account for only a small share of the achievement trends, however, because sizable direct effects would presum-

---

37. Lloyd D. Johnston, Patrick M. O'Malley, and Jerald G. Bachman, *Use of Licit and Illicit Drugs by America's High School Students, 1975-1984* (Washington, D.C.: National Institute on Drug Abuse, 1985), Table 10.

38. Lloyd Johnston, University of Michigan, personal communication, July 1985.

ably be limited to students who use particularly powerful or dangerous drugs or who use other drugs frequently. If alcohol use is excepted, such students constituted a small minority of the student population even during the years when drug use was at its peak. And while substantial alcohol use is reported by a large percentage of high school seniors, heavy--but less than daily--use of alcohol by seniors showed little change over time.39/

Television Viewing

Recent Nielsen estimates show average weekly television viewing of 28 hours by children ages 2 through 5; 27 hours by those ages 6 through 11; and about 22 hours by teenagers. 40/ Thus, the average 16-year-old has spent more time watching television than in formal schooling. The Advisory Panel on the Scholastic Aptitude Test Score Decline, voicing a common view, concluded that this extensive viewing contributed to the achievement decline and suggested that it did so:

- As a "thief of time," diverting children from other activities such as homework, reading, and writing that might contribute more to achievement;
- By raising expectations of entertainment that teachers cannot meet; and
- By fostering "simultaneous, visual, and affective" processes characteristic of the right hemisphere of the brain, rather than the linear, verbal, logical left-hemisphere processes that would contribute more to achievement test scores. 41/

Available data, however, contradict this assertion; whatever effects TV viewing has had on achievement more generally, it probably did not contribute to the aggregate trends in test scores of the past two decades.

---

39. Ibid., pp. 54, 56. For example, about half of male seniors reported at least one incident of drinking five or more drinks in a row in the last two weeks, but this figure has been fairly constant.

40. All estimates of viewing time in this section are based on data provided by the A. C. Nielsen Corporation.

41. Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination* (New York: The College Entrance Examination Board, 1977). While frequently voiced, the notion that television decreases both the attention span of young children and their patience for the slow pace and often uninteresting content of formal instruction has not been a focus of research and is not discussed here.

Many studies have found that high levels of television viewing are associated with lower levels of educational achievement. 42/ Whether television viewing actually causes lowered achievement, however, remains a matter of debate, although several recent studies suggest it has a negative effect on reading. This uncertainty stems in part from the difficulty of accounting for the effects of confounding variables--in this case, not only demographic and socioeconomic factors but also differences in intelligence-test scores of students watching dissimilar amounts of television. In addition, most American children watch a great deal of television, and this "restriction of range" in the amount of viewing probably attenuates estimates of television's effects. 43/

If the effects of television viewing on achievement are actually small, it might be because the activities from which television "steals time" are no more conducive to educational achievement than is TV viewing itself. The soundest studies of TV's effects on children's use of time, most of which unfortunately were conducted between 25 and 40 years ago and therefore reflect a much lower level of viewing than is currently the norm, suggest that the "activities most often replaced (by television) are those that can be considered functionally equivalent." When children increased their viewing of television, they reduced primarily the time they spent watching movies, listening to the radio, reading comic books, and playing with others. 44/ Among older children, very little time was taken from homework or reading books and magazines.

Whatever TV's effects on achievement in general, the timing of changes in the amount of viewing suggests that they were not an important influence on the aggregate achievement trends of the past two decades. Average viewing time in the mid-1970s was roughly comparable to that a decade earlier, at the onset of the decline (see Figure A-2). Viewing increased during the late 1970s, but the increases were larger in the younger age groups, among whom the achievement decline had already ended. The

---

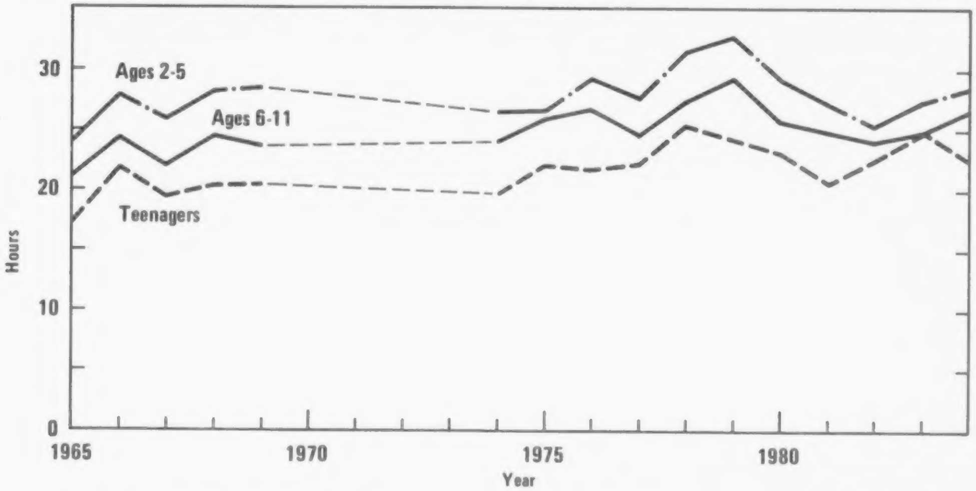
42. This section is based in large part on two recent reviews: Robert Hornick, "Out-of-School Television and Schooling: Hypotheses and Methods," *Review of Educational Research*, vol. 51 (Summer 1981), pp. 193-214; and Michael Morgan and Larry Gross, "Television and Educational Achievement and Aspiration," in David Pearl, Lorraine Bouthilet, and Joyce Lazar, eds., *Television and Behavior: Ten Years of Scientific Progress and Implications for the Eighties, Volume II--Technical Reviews* (Rockville, Md.: National Institute of Mental Health, 1982), pp. 78-90.

43. The severity of this last problem is indicated by one study that found that only two percent of students reported watching less than one hour per night.

44. Hornick, "Out-of-School Television and Schooling," p. 200.

Figure A-2.

Television Viewing by Children, by Age (Average hours per week)



SOURCE: Congressional Budget Office calculations based on A.C. Nielsen Co. unpublished data.

NOTE: Comparable data are not available for 1970 through 1973.

subsequent decline in viewing started at the end of the decade, at about the time that senior-high test scores started rising, but the drop lasted only a few years. Over the past several years, viewing has again increased, while tests scores at all grades have continued to climb.

Although the amount of viewing could not have contributed significantly to aggregate trends in test scores, changes in the content of the material viewed could nonetheless be germane. The available data do not permit assessing this hypothesis, however.

Students' Attitudes and Motivation

Changes in students' attitudes and motivation are among the most frequently cited explanations of the achievement decline. The Advisory Panel on the Scholastic Aptitude Test Score Decline, for example, referred to the late 1960s and early 1970s--the core years of the decline among high school seniors--as a "decade of distraction," a time when "national disillusionment" arose from the divisive Vietnam War, political corruption, assassinations, and large-scale urban riots. The Panel, noting that the students taking the SAT during the period of its sharpest decline had already experienced this social upheaval for five or six years and that some male students faced the prospect of the military draft after completing school, suggested that this may have negatively affected students' motivation and attitudes toward educational success. 45/ Other social trends of the period might reflect changes in attitudes that could also affect achievement--for example, the changes in drug use noted above, and the trends in suicide, homicide, and arrest rates among young people noted in Chapter IV.

This explanation appears to be among the most plausible. It fits many aspects of the achievement trends quite well, such as their timing; their remarkable pervasiveness among different types of students, schools, geographic areas, and subject areas; the existence of a subsequent upturn; and the fact that the decline was greater among older students. It is precisely the sort of broad societal change that could dramatically affect student achievement; at the same time, it is also consistent with many observers' accounts of the period.

On the other hand, this explanation seems impossible to test. In asserting the importance of attitudes and motivation, the Advisory Panel maintained that "the facts are as obvious as the proof of any causal relationship is impossible." Even that statement understates the difficulty of appraising the impact of these factors, for the "facts," however obvious, are hard to document systematically. Nationally representative surveys of students provide measures of attitudes and motivation, but the available data are both sparse and inconsistent. Between 1971 and 1979, for example, the proportion of high school seniors responding that their schools "should have placed more emphasis on basic academics" grew markedly--from roughly half to three-quarters. 46/ Some observers have taken this as a sign that students' interest in academic success remained high and that a decrease in the demands imposed by their schools made it harder to attain their desired level of achievement. Other results of the same survey, however, are inconsistent with this interpretation; for example, the proportion claiming that their courses were too hard increased from roughly 42 percent to 49 percent.

Lacking firsthand, systematic evidence, some analysts have used other educational trends as circumstantial evidence of relevant trends in students'

---

45. Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination* (New York: The College Entrance Examination Board, 1977), p. 37.

46. W. B. Fetters, G. H. Brown, and J. A. Owings, *High School Seniors: A Comparative Study of the Classes of 1972 and 1980* (Washington, D.C.: National Center for Education Statistics, undated), Table 2.7.

attitudes and motivation. Examples include grade inflation (an easing of the requirements for obtaining high grades) and the growth of "social promotion" (the promotion into higher grades of students who have not adequately mastered the material required in their present grades). It is unclear what these trends imply, however; they could be either effects or causes of relevant changes in students' attitudes and motivation, or they could be largely unrelated to them.

Trends in school attendance have been used as an indirect indicator of attitudes and might also be important as a measure of the total amount of schooling obtained. A major early review of the achievement decline argued that a drop in attendance was concurrent with the beginning of the decline and might have contributed to it. 47/ While changes in absenteeism are to some extent consistent with the achievement trends at the senior-high level, however, they have been slight. In the 1980 school year, for example, average daily attendance was 90 percent of school-year enrollment--the same as in 1959. Moreover, during this period, attendance was never more than 1.4 percentage points above or below 90 percent. 48/

Trends in the enrollment of senior-high students in academic and nonacademic programs might also reflect changes in student motivation and might be one of the mechanisms by which motivational trends affect test scores. In the senior class of 1972, 46 percent of all students were enrolled in academic programs; in the class of 1980, only 38 percent. Most of the corresponding increase was in "general" programs, although enrollments in vocational programs also grew a bit. While the cause of this change remains obscure, the fact that the shift out of the academic track was about twice as large among males as among females and that the relative growth in vocational enrollments only occurred among males suggests that these changes were in substantial part voluntary and, therefore, that students' attitudes and motivations might have played some role. 49/

Environmental Lead

Possible environmental explanations of the achievement trends have as yet generated relatively little attention, and information about them is typically

---

47. A. Harnischfeger and D. E. Wiley, *Achievement Test Score Decline: Do We Need to Worry?* (Chicago: CEMREL, Inc., 1975).

48. Center for Education Statistics, published and unpublished tabulations.

49. Fetters and others, *High School Seniors*, Table 2.1.

sparse. 50/ The research on the effects of environmental lead, summarized here, is unusually plentiful. It is important, not only because of the possible effects of lead itself, but because it illustrates the more general point that neither environmental factors nor students' health should be summarily ruled out as influences on trends in test scores.

The serious neurological effects of lead poisoning--which include gross impairment of both motor control and cognitive functioning, lethargy, convulsions, and even coma and death--have been documented for at least a century and a half. 51/ In addition, lead is widespread in the human environment because of lead-based paint, leaded gasoline, emissions from lead smelters, batteries, and other sources.

Existing research indicates that the exposure of many American children to lead has been sufficient to impair their cognitive functioning in ways that could affect performance in school. Individuals with levels of lead burden well below those that cause classic lead poisoning have shown lower scores on intelligence and other cognitive tests, poorer performance on perceptual-motor tasks, various disruptions of the functioning of the nervous system, and disturbances of attention. Children seem to be more susceptible to these effects than adults. Significantly, some of these problems are apparent in teachers' ratings of students with elevated levels of lead in their blood, suggesting that the symptoms appreciably interfere with students' functioning in school.

A considerable amount of the available research explores the effects of lead on performance on intelligence quotient (IQ) tests; the scores on these tests are highly correlated with those on many achievement tests. The research as a whole suggests that the IQ scores of children with notably elevated levels of lead in their blood (from 30 micrograms per deciliter to 70 micrograms per deciliter) but with no overt symptoms of lead poisoning appear to be depressed by 4 to 5 points--that is, by a fourth to a third of a standard deviation. Results of research about the effects of lesser exposure are less consistent, but some studies suggest a decrement of 1 to 2 points--0.07 to 0.13 standard deviation--at blood lead levels of 15 to 30

---

50. For a review of some environmental explanations, see B. Rimland and G. Larson, "The Manpower Quality Decline: An Ecological Perspective," *Armed Forces and Society*, vol. 8, no. 1 (Fall 1981), pp. 21-78.

51. A comprehensive review of the research on environmental lead, including a thorough discussion of methodological problems and gaps in existing data, can be found in Environmental Protection Agency, *Air Quality Criteria for Lead*, draft final version (Research Triangle Park, N.C.: Environmental Criteria and Assessment Office, June 1986).

micrograms per deciliter. Negative effects of lead on IQ are evident across the entire range of IQ scores. 52/

Although only a modest proportion of children have blood lead levels in the "notably elevated" range, many exceed 15 micrograms per deciliter and therefore may have appreciably depressed IQ scores. A national study in the late 1970s, for example--when exposure had already dropped considerably from earlier levels--found that 4 percent of children below age 6 had levels above 30 micrograms per deciliter, and almost 25 percent exceeded 20 micrograms per deciliter. Indeed, the average level among children under age 6 was about 15 micrograms per deciliter, and among older children it was about 12 micrograms per deciliter. 53/ Among certain groups--blacks, low-income children, and children in large metropolitan areas, for example--exposure is considerably greater yet. 54/

Recent data, though relatively sparse, consistently show a sharp decline in levels of lead in the blood that appears to reflect the reduction in the use of leaded gasoline. A large, nationally representative study found a drop in lead levels from 1976 to 1980 ranging from 31 percent to 42 percent. This reduction appeared in all age groups but was somewhat greater among children than adults. Other data from screening programs in individual cities show declines in lead levels of newborns and preschool children, in one case beginning as early as the late 1960s.

Although these drops in lead levels occurred in cohorts that produced rising test scores, gauging the temporal consistency of the two trends is difficult. One obstacle is that academic performance might be partly determined by past levels of lead exposure as well as current lead burden, because the effects of both lead exposure and education are cumulative. The absence of nationally representative data on lead burdens earlier than the mid-1970s is also problematic.

Although the effects of lead on the cognitive abilities of individual children can be large, any contribution of declining lead exposure to the aggregate rise in test scores would probably have been small. By way of

---

52. Ibid., vol. 1, p. 117, and vol. 4, pp. 12-86 and 12-95. Although early epidemiological studies of the effects of lead exposure have been criticized because of confounding with social class and other factors, recent reanalyses and studies appear to confirm that the relationships with cognitive functioning reported here are not an artifact of confounding.

53. Ibid., vol. 4, p. 11.16.

54. Ibid., vol. 4, pp. 11.15 and 11.20.

comparison, the IQ decrement of children with notably elevated levels of lead in their blood appears to be nearly as large as the average decline in achievement test scores in grades 6 through 12 and is larger than the increase to date shown by many tests. But, as noted earlier, few children have blood lead levels in the notably elevated range. Moreover, only changes in lead burden could have contributed to trends in test scores; stable, high levels of exposure could cause the average score to be lower than it would otherwise be but would not produce a change over time. Because lead exposure was substantial before the achievement decline and remains sizable today, the total change in lead exposure during the period in question was undoubtedly smaller than the highest levels of exposure reached during those years.

EDUCATIONAL FACTORS

Although educational changes have figured prominently in public discussion of the possible causes of the recent trends in test scores, systematic evidence supporting such a contribution is available for only a minority of the educational factors examined in this study. The available data contradict several common hypotheses and are simply inadequate to evaluate numerous others. 55/

Teachers' Skills and Experience

Few aspects of the educational system have been as central to the current debate as the quality of the teaching work force. Appraising the possible effects of teachers' characteristics on average test scores is impeded, however, by the remarkable inconsistency of much of the relevant cross-sectional research. The findings of many studies are statistically insignificant--that is, they might well be the result of chance. Furthermore, the few significant findings are often contradictory.

Teachers' Test Scores. Students intending to become teachers obtain, on average, relatively low scores on achievement tests, and their scores dropped more rapidly than those of students in general during the latter part

---

55. A number of educational factors that have figured prominently in the recent debate about achievement have been omitted from this section because they are not widely thought to have been causes of the specific test score trends analyzed in this study. Examples include the rise of real expenditures for education and the fall of pupil/staff ratios, both of which continued during the period of declining test scores.

of the achievement decline.[56] Moreover, higher-scoring teacher trainees are less likely to become teachers and, when they do, are more likely to leave the field for other occupations. The available data, however, do not indicate whether these latter two problems worsened during the period of the achievement decline. [57]

Cross-sectional research on the effects of teachers' academic achievement--as measured by tests such as the SAT and the ACT--on the achievement of their students is inconsistent. For example, one large-scale study of northern schools in the 1960s found that teachers' verbal abilities were consistently related to students' achievement.[58] In contrast, a large study of students in a single northeastern city found no relationship between teachers' scores on the National Teacher Examination and their students' test scores.[59] Nonetheless, it seems plausible that a deterioration of the academic skills of incoming teachers might in turn adversely affect the average test scores of their students.

Regardless of the cross-sectional research, however, the documented decline in the test scores of potential teachers, which began in the early 1970s, occurred too late to have contributed appreciably to the decline of students' test scores during the 1960s and 1970s. (The effects of any earlier deterioration of teachers' scores would be arguable, but, in any case, there are no data indicating whether one occurred.)

Most of the available data measure the academic achievement of a cohort of potential new teachers before they enter college--for example, the SAT and ACT scores of high school students planning to major in educa-

---

56. Theodore C. Wagenaar, "Occupational Aspirations and Intended Field of Study in College" (Washington, D.C.: National Center for Education Statistics, unpublished, November 1984); College Entrance Examination Board, *College Bound Seniors* (New York: CEEB, various years); American College Testing Program, *College Student Profiles: Norms for the ACT Assessment, 1983-84* (Iowa City: ACT Publications, 1983).

57. Victor S. Vance and Phillip C. Schlechty, "The Distribution of Academic Ability in the Teaching Force: Policy Implications," *Phi Delta Kappan*, vol. 64 (September 1982), pp. 22-27.

58. Eric Hanushek, "The Education of Negroes and Whites," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts (1968), cited in Henry M. Levin, "A Cost-Effectiveness Analysis of Teacher Selection," *Journal of Human Resources*, vol. 5 (1970), pp. 24-33.

59. Anita A. Summers and Barbara L. Wolfe, "Do Schools Make a Difference?" *American Economic Review*, vol. 67 (September 1977), pp. 639-652.

tion--rather than that of teachers themselves, and those students will not become teachers for at least five years, if at all. For example, a drop in the average SAT scores of students planning to become teachers between 1972 and 1973--the first years for which data are available--would have first affected the teaching work force in 1978, when the decline in test scores was essentially over. Even then, the impact on the average quality of the teaching work force would have been limited by the rate at which new teachers were hired, and the hiring of new teachers was restricted for much of the period of the achievement decline because of falling student enrollments. In 1970, for example, about 9 percent of all teachers in public schools were in their first year of teaching; that proportion had dropped to 5.5 percent in 1975 and to 1.6 percent in 1981.[60] If 5 percent of the teaching positions on average were filled by new teachers every year, the post-1973 cohorts with lower SATs would not have constituted a fourth of the teaching work force until at least 1982.

Beginning in 1980, the SAT scores of students expecting to major in education have risen appreciably (more rapidly than the scores of college-bound seniors as a whole). By the same logic, however, this change could not have contributed to the recent rise in students' test scores until 1985.

Teachers' Educational Attainment. The effects of teachers' educational attainment--that is, the highest level of education they have completed--on students' achievement remain controversial. (This specific question about the effects of teachers' educational attainment should not be confused with more general--and currently intensely controversial--questions about the value of pedagogical training.) One recent review, for example, noted that of 106 studies located, 95 showed no statistically significant effect of teachers' educational attainment. Of the remaining 11, about half found a positive relationship, and the remainder found a negative relationship.[61] Faced with these results, some researchers have concluded that teachers' educational attainment, at least beyond a bachelor's degree, has little or no effect on students' achievement, while others believe that methodological flaws underlie the absence of a relationship in many studies.

---

60. National Education Association, *Status of the American Public School Teacher, 1980-81* (Washington, D.C.: NEA, 1982), Table 6.

61. Eric A. Hanushek, "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, vol. 24 (September 1986), pp. 1141-1177.

Regardless of how one interprets this cross-sectional research, however, the educational attainment of practicing teachers offers no explanation for the test score decline because their educational attainment has risen without interruption for two decades. In 1960, about 15 percent of all public school teachers had less than a bachelor's degree, 62 percent had a bachelor's, and 23 percent had either a master's degree or six years of college-level education. In 1980, less than half a percent lacked a bachelor's degree. The proportion with only a bachelor's had dropped to about 50 percent, and those with either a master's degree (or six years) had risen to over 49 percent. 62/

Teachers' Experience. Trends in teachers' experience might have contributed to the recent rise in test scores but appear much less likely to have had any bearing on the previous decline. In both cases, however, the evidence is somewhat unclear.

Although cross-sectional research is quite inconsistent, it suggests that more experienced teachers may produce higher achievement in their students. One recent reviewer located 109 relevant studies; of the 40 that had statistically significant results, 33 showed higher achievement among students with more experienced teachers.63/ Another reviewer concluded that teachers' experience is related significantly to students' achievement only for the first five years of teaching. 64/

The mix of experienced and inexperienced teachers has changed substantially over the past two and a half decades, in part because trends in student enrollments altered the demand for new teachers. Assessing those changes, however, and gauging their temporal consistency with test score trends, are difficult. Experience can be measured in a number of ways, and trends in the various measures are not always consistent with each other. Changes in the proportion of teachers who are highly experienced, for example, do not keep pace with trends in the proportion who are novices. In addition, data on teachers' experience are drawn from a survey administered

---

62. All data used here on trends in teachers' educational attainment, experience, and attitudes are from National Education Association, *Status of the American Public School Teacher, 1980-81*.

63. Hanushek, "The Economics of Schooling." Studies with statistically insignificant results tended in the same direction, though less markedly.

64. Susan J. Rosenholtz, "Political Myths About Education Reform: Lessons from Research on Teaching," *Phi Delta Kappan*, vol. 66, no. 5 (January 1985), pp. 349-355.

only once every five years, which leaves the precise timing of these trends unclear. 65/

Teachers' average years of experience varied relatively little but was roughly consistent with the timing of achievement trends; it fell gradually from 13 years in 1960 to 10 years in 1975 and then rose again to 13 years in 1980. 66/ Only one of the more specific measures of changes in experience underlying those averages, however, lined up reasonably well with the timing of both the decline in test scores and the subsequent rise: the proportion of teachers with 20 or more years of experience. That proportion fell 28 percent in 1960 to 14 percent in 1975 and then rose to 22 percent in 1980. As noted earlier, however, there is some evidence that experience in excess of five years of teaching is not significantly related to students' achievement.

In contrast, the proportion of inexperienced teachers showed relatively little change during the period of declining test scores but fell sharply at roughly the time when test scores were rising. Between 1975 and 1980, the proportion of teachers with four or fewer years of experience dropped from 27 percent to 14 percent; the share with one or two years of experience fell from 11 percent to 5 percent; and the share with only one year of experience fell from 6 percent to 2 percent.

Other Characteristics of Teachers. The recent focus on the academic qualifications of teachers has perhaps obscured the question of whether there have been changes in other characteristics of teachers--such as their morale and attitudes toward students--that might influence student achievement. The limited data indicate that teachers' attitudes toward teaching have become increasingly negative over the past few decades. The proportion of public school teachers reporting that they would certainly become teachers again fell from about half in 1960 to 22 percent in 1980, while the proportion saying that they probably or certainly would not teach again grew from 11 percent to 36 percent.67/ Whether these changes in attitudes are related to the achievement decline, however--either as causes or as responses--cannot be determined.

---

65. All estimates of teachers' experience discussed here are taken from NEA, *Status of the American Public School Teacher, 1980-81*.

66. The median amount of experience varied less than the mean: it was 11 years in 1961; 8 years in 1966, 1971, and 1976; and 12 years in 1981.

67. Ibid., Tables 51 and 52.

State and Local Graduation Requirements

At least 41 states have raised their coursework requirements for graduation in the past several years. 68/ Part of the impetus for these changes was a widespread view that lax standards had contributed to the achievement decline. 69/

Regardless of whether state-mandated graduation requirements are too lax, they appear to have played no direct role in the achievement decline after 1974, for the requirements in effect in 1980 show remarkably little change from those of 1974. 70/ Systematic data are not available for the earlier years of the decline, however, and there are no comprehensive data on trends in requirements imposed by local districts. If local requirements were substantially lowered or if state requirements were eased before 1974, those changes might have influenced test scores, although their effects would probably have been largely limited to the higher grades, and they would not have affected students whose course loads exceeded even the early standard. If such undocumented changes actually occurred, they might have contributed to the greater severity or later end of the decline in the higher grades.

If they were sufficiently lax, however, graduation standards could have contributed indirectly to the decline in scores even if the requirements did not change. If large numbers of students were exceeding the requirements by a substantial margin before the decline, and if some other factor--changes in students' attitudes, for example--caused them to seek an easier course load, lax standards would permit coursework to decline more than would be possible in the presence of stricter standards. To gauge whether graduation requirements had this sort of indirect role, it is necessary to consider changes in the actual coursework of students, which are assessed below.

---

68. Margaret E. Goertz, *State Educational Standards: A 50-State Survey* (Princeton: Educational Testing Service, January 1986); see also "Changing Course: A 50-State Survey of Reform Measures," *Education Week*, vol. 4 (February 6, 1985), pp. 11-30.

69. See, for example, National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform* (Washington, D.C.: Department of Education, April 1983), pp. 18, 20.

70. National Association of Secondary School Principals, *Graduation Requirements* and *State-Mandated Graduation Requirements, 1980* (Washington, D.C.: NASSP, 1975 and 1980). If standards were too lax throughout this period, the achievement of certain students might have been lower as a result. Only modification of those requirements, however, not lax but stable requirements, could cause average scores to change.

Coursework

Evidence strongly suggests that changes in coursework at the secondary school level were substantial and could have contributed appreciably to the achievement decline in those grades. It is necessary, however, to distinguish the number of courses taken from the content and difficulty of those courses. Changes in the number of courses taken are germane primarily to students in grades 7 and above. Changes in course content and difficulty, on the other hand, can occur at all grade levels. While information about trends in elementary school coursework is lacking, there is evidence about both the number and content of secondary school courses. 71/

That students scoring higher on achievement tests typically have taken more of the relevant courses is well established. Two nationally representative studies, for example, have found a sizable association between the amount of coursework in mathematics and mathematics test scores.72/ This association between coursework and test scores undoubtedly exaggerates the effects of coursework to some degree, because students who take more courses in difficult subjects have other characteristics--such as greater aptitude, prior achievement, and motivation--that would also contribute to their higher test scores. Nonetheless, research suggests that the independent effect of coursework on test scores is large. 73/

---

71. To the extent that textbooks both reflect and shape instruction, changes in texts--which are discussed in the following section--can also be seen as an indication of trends in course content.

72. Wayne W. Welch, Ronald E. Anderson, and Linda J. Harris, "The Effects of Schooling on Mathematics Achievement," *American Educational Research Journal*, vol. 19 (Spring 1982), pp. 145-153; and Lyle V. Jones, "White-Black Achievement Differences," *American Psychologist*, vol. 3, no. 11 (November 1984), pp. 1207-1213.

73. One recent, nationally representative study of seniors found that controlling for socioeconomic status, race, sex, and sophomore-year mathematics achievement left a large--though weakened--association between the number of core mathematics courses taken in high school and senior-year mathematics achievement. (Marshall Smith, Stanford University, unpublished analysis of the High School and Beyond survey.) Even after adjustment for these variables, the difference between the two extreme groups--those taking no mathematics courses and those taking algebra 1 and 2, geometry, and trigonometry-- was nearly a full standard deviation. A portion of that remaining gap in test scores, however, might in part reflect still other confounding factors, such as motivational disparities among students choosing to take different numbers of mathematics courses.

Simple tabulations of the average number of courses taken by secondary school students, without regard to the content or difficulty of the courses, give varying results and do not consistently parallel achievement trends. For example, tabulations of the average number of courses per student in grades 7-12 of public schools show that enrollments in English and language arts courses increased slightly between the late 1940s and early 1970s; mathematics courses increased through 1970 (through the first half of the achievement decline), then dropped substantially over the next two years; and science enrollments rose slightly from 1960 to the early 1970s. 74/ Data for seniors in the classes of 1972 and 1980 (a period covering the latter part of the decline and the first few years of the upturn in that grade) show a 10 percent increase in the average number of semesters of mathematics completed but a 13 percent decline in social studies and a 21 percent decline in foreign languages. Changes in English and science were trivial. 75/

The difficulty of the courses taken apparently decreased substantially, however, at least during the later half of the achievement decline. One indication of this is the marked shift out of academic programs noted earlier. A second indication is that the proportion of coursework devoted to remedial study soared. Between the 1971 and 1979 school years, for example, the proportion of seniors who had taken remedial mathematics grew more than sevenfold, from 4 percent to 30 percent. The proportion taking remedial English courses grew similarly. 76/ Finally, enrollment in so-called fringe courses, such as science fiction, appears to have grown rapidly, and there is some evidence that enrollment in such courses--presumably, as a substitute for such core subjects as English composition--was associated with greater declines in SAT scores. 77/

---

74. A. Harnischfeger and D. E. Wiley, *Achievement Test Score Decline: Do We Need to Worry?* (Chicago: CEMREL, Inc., 1975), Table 13.

75. W. B. Fetters, G. H. Brown, and J. A. Owings, *High School Seniors: A Comparative Study of the Classes of 1972 and 1980* (Washington, D.C.: National Center for Education Statistics, undated), Table 2.3.

76. Ibid.

77. The evidence pertaining to fringe course enrollments and their effects is not nationally representative but is nonetheless substantial. See Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination* (New York: College Entrance Examination Board, 1977).

The narrowing of the achievement gap between black and nonminority students was also paralleled by slight increases in the number of courses taken by black high school students, relative to the courses taken by nonminority students. 78/ This increase in coursework, however, is probably in part an effect of other factors contributing to the narrowing achievement gap between black and nonminority students, rather than solely its cause. That gap in achievement has narrowed similarly in the lower grades--a pattern that cannot be attributed to course enrollments as such but that could reflect changes in the educational experiences of black students that are also manifested in changing senior-high course enrollments. 79/

Minimum-Competency Testing

Some analysts who attribute part of the achievement decline to loosened educational standards also attribute the recent upturn to the growth of minimum-competency (or "competency" or "mastery") testing, which they see as part of a return to greater accountability and tougher standards. 80/

Although the effects of competency testing remain a matter of vehement debate, the timing of the upturn in achievement indicates that the growth of competency testing, whatever its effects on test scores generally, did not help initiate the rise in scores. Most of the increase in state-mandated competency testing occurred in the late 1970s--that is, several years after the upturn in achievement first became apparent in the lower grades. Fewer than a third of the states had even mandated (let alone implemented) competency-testing programs by the end of 1976, by which time the upturn in achievement had already been under way a few years and

---

78. Donald A. Rock, Ruth B. Eckstrom, Margaret E. Goertz, Thomas L. Hilton, and Judith Pollack, *Factors Associated with Decline of Test Scores of High School Seniors, 1972 and 1980* (Washington, D.C.: Center for Statistics, Department of Education, December 1985), Tables 6-48 through 6-52. In some instances, the number of relevant courses taken by black students increased more than the number taken by nonminority students; in others, black students showed a smaller decrease.

79. The data noted here do not address possible increases in the difficulty of the courses taken by black students, which--if they occurred at all--could have affected younger students as well.

80. See, for example, Barbara Lerner, "The Minimum Competency Testing Movement: Social, Scientific, and Legal Implications," *American Psychologist*, vol. 36 (October 1981), pp. 1057-1066.

had reached approximately grade 8. 81/ Moreover, one would expect some delay before the effects of the new testing programs would have been fully manifested.

Textbook Difficulty

The role that changes in texts might have played in recent achievement trends remains unclear. Anecdotal reports of the "dumbing down" of textbooks are so widespread that they should not be discounted, but systematic evidence remains very sparse and is not entirely consistent with achievement trends. Moreover, cross-sectional data evaluating the impact that relevant changes in texts might have had on achievement is lacking.

Reading, Language Arts, and History. Most references to systematic data on the difficulty of textbooks reflect a single study commissioned by the Advisory Panel on the Scholastic Aptitude Test Score Decline. 82/ This study examined texts in reading and literature, grammar and composition, and history at the first-, sixth-, and eleventh-grade levels published over a period of about five decades, beginning with the 1920s. It considered many aspects of text difficulty, including sentence length; vocabulary difficulty; demands for reading, writing, and reasoning in assignments; the degree of "child-centeredness" of the material; and the organization and coherence of the text.

While a number of striking changes were found, they varied markedly among grade levels and subject areas, and their timing was not entirely consistent with that of achievement trends. 83/ For example, first-grade

---

81. One recent review offered the following chronology: "...It appears that only two states had mandated minimum competency testing programs as early as 1971, that four more took similar actions in 1972, but that none were added during 1973 and 1974. In 1975, five more states enacted minimum competency testing programs, and four more were added in 1976....The rapid acceleration of the movement can be noted from data for 1977, when an additional nine states mandated...programs." (R. M. Jaeger, "The Final Hurdle: Minimum Competency Achievement Testing," in G. R. Austin and H. Garber, eds., *The Rise and Fall of National Test Scores* (New York: Academic Press, 1982), p. 228.)

82. J. S. Chall, S. S. Conard, and S. H. Harris, *An Analysis of Textbooks in Relation to Declining SAT Scores* (New York: College Entrance Examination Board, June 1977).

83. This partial consistency might explain why this study is cited both by those arguing that change in textbooks has been substantial (for example, Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination*) and by those maintaining that change has been minor (for example, Christopher Jencks, "Declining Test Scores: An Assessment of Six Alternative Explanations," *Sociological Spectrum*, Premier Issue (December 1980), pp. 1-15.

readers were already becoming easier by the 1930s, and this trend continued at least until 1956 and, by some measures, for half a decade after that.[84] Since texts are used by students who take the SAT roughly 10 to 15 years after the publication date, progressively easier reading texts were used by both the cohorts responsible for the sharp rise in SAT scores in the late 1950s and early 1960s (that is, before the decline) and those that produced the first part of the decline.

Trends in sixth-grade reading texts showed even less consistency with SAT trends. During most of the period when the cohorts responsible for the SAT decline were attending sixth grade, basic readers were stable or increasing in difficulty. Cohorts taking the SAT in the late 1970s, whose scores marked the lowest point in recent years, used texts that were at least as difficult as those used by cohorts that took the SAT before the scores began to decline.[85] Trends in the difficulty of exercise questions included in the texts also failed to conform consistently to trends in the SAT. For example, the level of cognitive difficulty of questions in sixth-grade history texts was higher in texts published from 1965 through 1970--that is, in texts used by the cohorts that produced the second half of the SAT decline--than in texts published in the 1950s.[86]

The results of that study are also inconsistent with one of the most pervasive aspects of recent achievement trends: the greater severity of the decline in the higher grades. To the extent that a measurable lessening of reading difficulty was found, it tended to be more pronounced in the earliest grades, and the changes in the first-grade texts came closest to paralleling the timing of test score trends. One could easily posit a process by which changes in instruction in the earlier grades affect achievement in the higher grades as well. It is more difficult, however, to conceive of instructional changes that would have no impact for several years--there was no achievement decline in the first three grades--but would have progressively larger effects thereafter.[87]

---

84. Chall and others, *An Analysis of Textbooks*, Table 5.

85. Ibid., Table 6.

86. Ibid., Table 17.

87. The principal author of that study hypothesized that the achievement decline in the higher grades could indeed be explained by long-term effects of early reading instruction. See also Jeanne S. Chall, "Literacy: Trends and Explanations," *Educational Researcher*, vol. 12, no. 9 (November 1983), pp. 3-8.

Mathematics. The 1960s and 1970s saw some dramatic alterations to elementary mathematics texts that began with the incorporation of the "new math" in texts published between 1963 and 1965. (The new math had begun earlier but was not reflected in a major text until 1963.) The new texts also included fewer and simpler word problems, beginning a trend that continued for about a decade and a half. 88/ These changes are consistent with the timing of the achievement decline and fit with the greater drop in performance in story problems and other applications compared with simple computation.

Texts used in the early 1980s and those recently introduced, which will be the standards for the latter half of the 1980s, show a marked increase both in the number of word problems and in the proportion of problems comprising more than a single step. Mathematics achievement scores began rising in the 1970s, however. In addition, the National Assessment found that during the time of the upturn, students improved least in the area of applications--a pattern that would not be expected if students spent more time practicing the solution of multistep word problems.

Although the return of word problems in mathematics textbooks thus could not have initiated the upturn in mathematics test scores, it might reflect a return to "old" mathematics that was already under way and might have reinforced whatever effects the prior change in course content had on achievement. Moreover, the latest National Assessment of mathematics was conducted in the 1981-1982 school year, and the effects of the most recent changes in texts might appear as increases in problem-solving ability only in subsequent assessments.

Homework

Time spent on homework dropped during the latter years of the achievement decline--at least among seniors--and has risen again recently. These changes might have contributed to trends in scores on achievement tests, but their effects were probably small.

---

88. Susan R. Stockdale, "An Analysis of Elementary Mathematics Textbook Story Problems During the Eighties, and Comparisons to Earlier Eras" (doctoral dissertation, University of Iowa, Iowa City, May 1985); H. D. Hoover, Iowa Testing Programs, personal communication, October 1985.

Many cross-sectional studies show an association between achievement and the amount of homework done.[89] The extent to which this association reflects homework itself and not other factors (such as higher motivation, more substantial prior coursework, or the higher ability of students who do more homework) has not been fully determined, but it is reasonable to expect that, up to a point, increases in homework can raise test scores. [90]

According to reports by high school seniors, however, the decline in the average time spent on homework during the 1970s was small, and the average was already low. Between 1971 and 1979, the average time seniors reported spending on homework dropped from 4.3 to 3.9 hours per week--a decline of roughly 25 minutes a week. Oddly, the proportion of seniors doing no homework dropped, and the share doing more than 10 hours per week increased a bit (see Table A-1). [91] Between the 1979 and 1983 school years, both the amount of time spent on homework by 17-year-old students and the proportion of students assigned homework the previous day rose, but those changes too were small. [92]

---

89. This association was found, for example, in Rock and others, *Factors Associated With Decline of Test Scores of High School Seniors, 1972 and 1980* (Washington, D.C.: Center for Statistics, Department of Education, December 1985), and in Timothy Z. Keith, "Time Spent on Homework and High School Grades: A Large-Sample Path Analysis," *Journal of Educational Psychology*, vol. 74, no. 2 (1982), pp. 248-253. In the most recent National Assessment of reading, on the other hand, the amount of homework done was unambiguously associated with achievement only in the oldest sample (age 17); see Bernice Anderson, Nancy Mead, and Susan Sullivan, "Homework: What Do National Assessment Results Tell Us?" (Princeton: NAEP/Educational Testing Service, December 1986).

90. Keith ("Time Spent on Homework and High School Grades") used nationally representative data to disentangle some of these confounding factors from the association between homework and class grades, but he did not control for prior coursework and controlled only inadequately for students' ability and prior level of achievement, because of limitations of the data.

91. W. B. Fetters, G. H. Brown, and J. A. Owings, *High School Seniors: A Comparative Study of the Classes of 1972 and 1980* (Washington, D.C.: National Center for Education Statistics, undated).

92. Anderson and others, "Homework: What Do National Assessment Results Tell Us?"

While the direct effect of this small change in homework was probably small, it might also have had an indirect bearing on achievement trends. The amount of homework might be a reflection of other correlated trends, such as changes in teachers' expectations and students' motivation, that in turn directly affect achievement.

Demands for Writing

Changes in schools' demands for writing have figured prominently in the recent debate--both as a possible cause of the achievement decline and as a component of efforts to improve education--and some states have recently added writing assessments to their battery of mandatory competency tests.

TABLE A-1. PERCENT OF HIGH SCHOOL SENIORS REPORTING VARIOUS AMOUNTS OF TIME SPENT ON HOMEWORK PER WEEK, 1971 AND 1979 SCHOOL YEARS

Amount of Time	1971	1979
None	11	8
Under 5 Hours	54	68
Five to 10 Hours	30	18
Over 10 Hours	6	6
Average Number of Hours Per Week	4.3	3.9

SOURCE: William Fetters, G. H. Brown, and J. A. Owings, *High School Seniors: A Comparative Study of the Classes of 1972 and 1980* (Washington, D.C.: National Center for Education Statistics, undated), Table 2.7.

Although anecdotal reports of a declining emphasis on writing are too widespread to dismiss out of hand, there appears to be little systematic evidence of such a decline. The Advisory Panel on the Scholastic Aptitude Test Score Decline concluded that demands for writing did fall, but it based that conclusion in part on the study of textbooks cited earlier, and the trends in demands for writing found in that study were not entirely consistent with achievement trends. For instance, the type and amount of writing required by sixth-grade reading texts and grammar and composition books were stable over the years studied.[93] Similarly, a comparison of the frequency with which seniors reported being assigned to write essays, themes, poetry, or stories changed only trivially between 1971 and 1979.[94]

As in the case of graduation standards, some of the apparent inconsistency between anecdotal and systematic information might represent the difference between low, but stable, demands and declining demands. Most observers would agree that the writing abilities of many American students need improvement. Yet trend data on actual writing--as opposed to proxies such as multiple-choice tests of English usage--are extremely sparse and do not paint a clear picture of declining performance.[95] Similarly, the fact that data on changes in demands for writing are limited and do not clearly show a decline in standards does not at all imply that demands for writing are adequate.

Grade Inflation

Many observers have cited "grade inflation"--the lowering of the level of achievement required to obtain a given grade--as a symptom of the lessened demands of schooling during the last few decades and maintain that this trend contributed to the decline in achievement.

---

[93] Chall and others, *An Analysis of Textbooks*, pp. 25, 32, and 33. In the case of grammar and composition books, the study reported a drop in the proportion of assignments requiring the writing of paragraphs, themes, stories, and so on (as opposed to single sentences and the like). This apparent trend, however, turns out to represent the texts of a single publisher; consideration of a single later text from a second publisher leads to the opposite conclusion (ibid., Table 13).

[94] Rock and others, *Factors Associated with Decline of Test Scores*, Table 4-25.

[95] For the most recent data on writing abilities, see National Assessment of Educational Progress, *The Writing Report Card* (Princeton: NAEP/Educational Testing Service, 1986). The most recent NAEP results are not comparable, however, to earlier data. For data on changes in writing performance, see National Assessment of Educational Progress, *Writing Achievement, 1969-1979* (Denver: NAEP/Education Commission of the States, 1980).

Although grade inflation at the secondary school level was clearly substantial and indicated declining educational demands, its link to the decline in test scores remains only speculative. 96/ Available data are not adequate to appraise fully the consistency of this change in educational practice with the timing or details of the achievement trends. The data do, however, indicate one inconsistency between grading patterns and achievement trends that might suggest that the impact of the former on the latter was small: seniors in Catholic schools experienced no appreciable grade inflation but nonetheless showed declines in test scores roughly comparable to those of public school students. 97/

Educational Programs for Disadvantaged Students

Over the past several years, a number of observers have suggested that federally funded educational programs for disadvantaged students contributed to certain aspects of the achievement trends. The Title I (now Chapter 1) compensatory education program has been noted most often in this regard, but Head Start has also been mentioned. 98/ This section

---

96. For data on grade inflation, see Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination*, p. 29; and National Center for Education Statistics, *The Condition of Education: 1982 Edition* (Washington, D.C.: Department of Education), p. 76.

97. Ibid.; Rock and others, *Factors Associated with Decline of Test Scores*, Tables 5-1 through 5-3. Data on students in non-Catholic private schools were too scanty to draw meaningful conclusions.

98. Statement by Archie E. Lapointe, Executive Director of the National Assessment of Educational Progress, before the House Subcommittee on Elementary, Secondary, and Vocational Education, Committee on Education and Labor, January 31, 1984; Archie E. Lapointe, "The Good News About American Education," *Phi Delta Kappan*, vol. 65 (June 1984), pp. 663-668; National Assessment of Educational Progress, *The Reading Report Card: Progress Toward Excellence in Our Schools* (Princeton: NAEP/ Educational Testing Service, 1985); National Assessment of Educational Progress, *Reading, Science and Mathematics Trends: A Closer Look* (Denver: NAEP/ Education Commission of the States, December 1982); National Assessment of Educational Progress, *Has Title I Improved Education for Disadvantaged Students? Evidence from Three National Assessments of Reading* (Denver: NAEP/ Education Commission of the States, April 1981).

assesses the contributions of these two programs to the relative gains of some minority groups and to the relatively favorable trends among younger children.

Head Start. Although existing research suggests that preschool programs for disadvantaged students can have diverse benefits, it does not indicate that Head Start contributed appreciably to the relative gains of black and Hispanic students observed in data on national achievement tests.

Over the past decade, a number of research reports have indicated that preschool programs for disadvantaged children can have lasting effects on school success. 99/ For example, students in some programs are less likely to be placed in special education or to repeat subsequent grades in school. While these reports have led many observers to express optimism about the effects of Head Start, the extent to which the effects of those preschool programs indicate comparable effects of Head Start is a matter of debate. Few of the programs were actually Head Start programs; most were experimental programs run by researchers and probably differed significantly from the typical Head Start programs of the time. Some programs were small and thus provide only a weak basis for inferences about the nation as a whole; indeed, one of the most prominent studies--the Perry Preschool Project--included only 58 children in the experimental groups.

---

99. This discussion of the general effects of Head Start and other preschool programs reflects the following studies: John R. Berrueta-Clement, Lawrence J. Schweinhart, W. Steven Barnett, Ann S. Epstein, and David P. Weikart, *Changed Lives: The Effects of the Perry Preschool Program on Youths Through Age 19* (Ypsilanti, Mich.: High/Scope Press, 1984); Urie Bronfenbrenner, *A Report on Longitudinal Evaluations of Preschool Programs, Volume II: Is Early Intervention Effective?* (Washington, D.C.: Department of Health, Education, and Welfare, 1976); Consortium for Longitudinal Studies, *Lasting Effects After Preschool* (Washington, D.C.: Department of Health, Education, and Welfare, 1979); Richard B. Darlington, Jacqueline M. Royce, Ann Stanton Snipper, Harry W. Murray, and Irving Lazar, "Preschool Programs and Later School Competence of Children from Low-Income Families," *Science*, vol. 208 (April 11, 1980), pp. 202-204; Head Start Evaluation, Synthesis and Utilization Project, *The Impact of Head Start on Children, Families and Communities* (Washington, D. C.: CSR, Incorporated, March 1985); Irving Lazar, Richard Darlington, Harry Murray, Jacqueline Royce, and Ann Snipper, "Lasting Effects of Early Education," *Monographs of the Society for Research in Child Development*, vol. 47, nos. 2-3, Serial no. 195 (1982); and New York State Education Department, *Evaluation of the New York State Experimental Prekindergarten Program* (Albany: The University of the State of New York, February 1982). The specific effects of Head Start on ethnic disparities in test scores are not assessed in those studies, however.

Evaluations of Head Start programs themselves do not fully clarify the extent to which their effects have been comparable. 100/

In addition, the documented long-term benefits of preschool programs generally have not included higher test scores. 101/ A great many programs--including some Head Start programs--have shown short-term improvement in intelligence (IQ) test scores, but these gains typically are largely or entirely eroded within several years. Assessments of effects on achievement test scores are much less common, and most show a similarly rapid erosion of gains. Evidence of these gains lasting into the secondary school years is very sparse and is found mostly in the smaller experimental programs. 102/

Moreover, even if Head Start raised achievement test scores, its contribution to the relative gains of minority students would be constrained by the substantial participation of nonminority students in the program. While black and Hispanic students are indeed overrepresented in the program relative to their numbers in the cohort as a whole, non-Hispanic whites nonetheless account for about a third of enrollments. In addition, relatively few children participate in the program, which further dilutes any effects on aggregate test scores. Head Start enrollments have generally ranged between 3 percent and 5 percent of children ages three to five. 103/

Given these considerations, it would be reasonable to expect Head Start to have raised the aggregate test scores of black and Hispanic students relative to those of nonminority students by perhaps 0.02 standard deviation one year after participation in the program, and perhaps by half or two-thirds that much two years after. (By contrast, the relative gains of

---

100. Studies of Head Start's effects are numerous, but many are seriously flawed, and arguments about biases in even the more substantial evaluations are common. In addition, few studies of Head Start have followed students for the length of time required to assess long-term influences on school performance.

101. The long-term effects that have received the greatest attention have been changes in other aspects of school success, such as changes in rates of assignment to special education.

102. Berrueta-Clement and others, *Changed Lives;* Head Start Evaluation, Synthesis and Utilization Project, *The Impact of Head Start;* Lazar and others, "Lasting Effects of Early Education."

103. Administration for Children, Youth, and Families, *Project Head Start Statistical Fact Sheet* (Washington, D. C.: Department of Health and Human Services, December 1985); Department of Commerce, Bureau of the Census, *Estimates of the Population of the United States by Age Sex and Race,* Series P-25, nos. 519, 917, and 1,000 (various years).

black students on the SAT were 0.16 and 0.21 standard deviation, depending on the subtest.) By the third year, however--that is, by the earliest grade reflected in the test score data discussed here--any effects of the program would be negligible.

Head Start could have contributed to the relatively favorable trends among younger students, but this effect would have been trivial because few children in each cohort attended the program. It is plausible, for example, that the program raised scores in the earliest grades by 0.01 standard deviation or less relative to scores in higher grades, but an effect of that size is negligible compared with the total difference in trends among age groups.

Title I/Chapter 1 Compensatory Education. Evaluations of the Title I/Chapter 1 program have consistently shown that the program has a small effect on achievement test scores. The evidence as a whole suggests that:

- Gains in test scores of students in the program exceed those of comparable students not in the program by roughly 10 percent to 30 percent, depending on age and subject;
- These gains are not large enough to narrow substantially the gap between program participants and other students;
- The program's impact is greater in mathematics than in reading and larger in the lower grades than in the higher grades; and
- The gains of participating students erode after students leave the program. 104/

In terms of its possible effects on disparities in test scores among ethnic groups, Title I/Chapter 1 differs from Head Start in three important respects. First, because students participating in the program are of school age and thus contribute to aggregate test scores while in the program, even transitory effects of Title I/Chapter 1 could narrow the gap between ethnic groups. In addition, Title I/Chapter 1 is a much larger program than Head Start--currently, about 14 percent of all students in kindergarten through grade 8 participate in the program--so any impact on the test scores of participating students has that much larger an effect on aggregate test

---

104. For a current overview of the copious research evaluating this program, see National Assessment of Chapter 1, *The Effectiveness of Chapter 1 Services* (Washington, D.C.: Department of Education, 1986).

scores. Finally, the ethnic composition of the group of students served by Title I/Chapter 1 is more similar to that of the student body as a whole than is the group served by Head Start; 45 percent of Chapter 1 students in 1984, and 54 percent of those served in the mid-1970s, were non-Hispanic whites (compared with 73 percent to 78 percent of the school-age population in those years). 105/ This greater similarity would ameliorate any effect of the program on disparities in test scores among ethnic groups.

Thus, Title I/Chapter 1 could have contributed measurably to the relative gains of black and Hispanic students, but probably only in the early grades. Although a precise estimate is impossible, Title I/Chapter 1 narrowed the gap between black and nonminority students in grade 4 by roughly 0.04 to 0.06 standard deviation and that between Hispanic and nonminority students by 0.02 to 0.05 standard deviation--a small effect in absolute terms, but a moderate share of the total relative gains of those groups. 106/ In the higher grades, however, the effect of the program would have been far smaller--perhaps even negligible--because of the much smaller percentage of students participating in the program in those grades, the lesser impact of the program on older students, and the apparent lack of persistence of effects on younger program participants.

Finally, Title I/Chapter 1 could have made a minor contribution to the relatively favorable trends in the youngest children because of the small number of students served by the program in the higher grades. It is plausible, for example, that Title I/Chapter 1 might have raised aggregate scores in the first four grades by roughly 0.025 standard deviation relative to scores in the twelfth grade. Such an effect, however, would be an order of magnitude smaller than the observed difference in trends among the youngest and oldest children.

Desegregation

Some observers have suggested that the relative gains of certain minority students might in part reflect the effects of desegregation. To evaluate this

---

105. The 1984 Chapter 1 estimate reflects data provided by the Department of Education. The mid-1970s Title I estimate is from National Institute of Education, *Compensatory Education Services* (Washington, D.C.: Department of Health, Education, and Welfare, July 1977), Table 2. Estimates of the composition of the school-age population are based on CBO tabulations of the Current Population Survey.

106. For present purposes, the most important weakness of the existing data is the absence of research distinguishing the program's effects among different ethnic groups. These estimates assume that the effect of the program is the same regardless of ethnicity.

hypothesis, it is necessary to distinguish between the relative gains of black and Hispanic students because of different trends in segregation experienced by the two groups.

Black Students. Research on the effects of desegregation on the test scores of black students is plentiful but inconsistent. Some inconsistency should probably be expected; the effects of desegregation presumably would differ among locations, depending on the characteristics of the communities, schools, nonminority students, and black students involved.

A recent synthesis of research concluded that, in the aggregate, desegregation probably increased the reading scores of black students. Quantifying that gain proved difficult, however, because the estimate varied greatly depending on the technical criteria used to decide which studies were of sufficiently high quality to be credible. The review concluded that the gains of directly affected black students were probably in the range of 0.06 to 0.16 standard deviation, although some studies suggested that the upper bound of the estimate should be higher--about 0.26 standard deviation. In mathematics, on the other hand, the effect of desegregation, if any, appeared trivial. 107/

The contribution of desegregation to the relative gains of black students in the aggregate, however, would have been considerably smaller than the gains of directly affected students, because for many black students the amount of segregation experienced did not change markedly. That is, even though the amount of desegregation between the late 1960s and the present has been substantial, a sizable share of black students remain in segregated environments. In addition, some black students were in desegregated environments before desegregation began.

The results of research on desegregation are also somewhat inconsistent with the observed relative gains of black students in the aggregate. Those gains were not limited to reading; indeed, some tests showed greater

---

107. Thomas D. Cook, "What Have Black Children Gained Academically From School Integration?: Examination of the Meta-Analytic Evidence," in Thomas D. Cook, David Armor, Robert Crain, Norman Miller, Walter Stephan, Herbert Walberg, and Paul Wortman, *School Desegregation and Black Achievement* (Washington, D.C.: Department of Education, May 1984). This article reviewed and synthesized the results of a number of other syntheses of individual studies and examined the factors that might account for varying estimates among reviews.

gains in mathematics. This finding suggests that if the research is correct in showing at most a trivial effect of desegregation on achievement in mathematics, desegregation's direct effects leave much of the relative gains of black students unexplained.

Gauging the temporal consistency of desegregation and the relative achievement gains of black students is problematic, for it is not apparent what point in students' school careers to align with changes in segregation. Nationally, desegregation occurred primarily before 1971 or 1972. The proportion of black students attending predominantly minority schools (schools with minority enrollments over 50 percent) declined from 77 percent in 1968 to 64 percent in 1972. In contrast, the decline over the next eight years was negligible--to 63 percent. The proportion of black students attending schools with minority enrollments of 90 percent or more showed a similar trajectory, declining from 64 percent in 1968 to 39 percent in 1972 and 33 percent in 1980. 108/ A similar pattern emerged in a study of the degree of within-district segregation of black and nonminority students in 116 central-city school districts. 109/

Thus, little desegregation occurred during the years when black students were gaining on achievement tests relative to their nonminority peers. But if segregation in the early years of schooling is especially important, trends in desegregation would nonetheless be temporally consistent with some of the relative achievement gains of black students. For example, the most recent analysis of NAEP reading trends shows that the largest relative gains of black students occurred among those in the cohorts born roughly between 1961 and 1967--the cohorts that entered school during

---

108. Gary Orfield, *Working Paper: Desegregation of Black and Hispanic Students From 1968 to 1980* (Washington, D. C.: Joint Center for Political Studies, 1982), Table 11.

109. Reynolds Farley, "Trends in School Segregation and Enrollment by Race: An Analysis of New Data From the Office of Civil Rights" (University of Michigan Population Studies Center, Ann Arbor, unpublished final report to the National Institute of Education, October 1981). This report measured segregation differently: it assessed disparities in the racial mix of schools within a district. By this measure, a district in which all minority students attended schools with a high percentage of minority students would nonetheless be considered fully desegregated if the ethnic mix was identical in all schools. This measure, which is entirely insensitive to changes in the composition of the student body resulting from factors such as declining nonminority enrollments, is relevant in many legal contexts. When considering the effects of desegregation on achievement, however, it is likely that the relevant indices are those--such as Orfield's--that measure the composition of the schools attended by minority students, regardless of the causes of that composition.

the years of most rapid desegregation. 110/ Similarly, the relative gains of students on the SAT were apparent by the time of the cohort that entered first grade in 1965 and appear to have ended after the cohorts that entered in 1970 and 1971. On the other hand, statewide data from North Carolina and Texas show relative gains by black students in later cohorts.

Hispanic Students. The relative achievement gains of Hispanic students, unlike those of black students, clearly did not stem from desegregation. During the period for which data are available, Hispanic students became more segregated, not less. In 1968, 55 percent of Hispanic students attended schools with predominantly minority enrollments; 23 percent were in schools with minority enrollments of 90 percent or more. By 1980, those proportions had risen to 68 percent and 29 percent, respectively. 111/

SELECTION FACTORS

This section explores two types of selection changes: trends in the proportion of students remaining enrolled in school (retention) and changes in the proportion of students choosing to take college admissions tests (self-selection). Several other aspects of selection, such as changes in the enrollment of certain types of handicapped students and in the policies governing the participation of handicapped students or students with limited proficiency in English in routine testing programs, are not discussed because appropriate data are not available. 112/

---

110. National Assessment of Educational Progress, *The Reading Report Card*, Figure 3.1.

111. Orfield, *Desegregation of Black and Hispanic Students from 1968 to 1980*. Similarly, from 1970 to 1980, the average minority enrollment in schools attended by Hispanic students rose from 56 percent to 64 percent. Another study reported a modest decline in the segregation of Hispanic students during the late 1960s and early 1970s in 44 central city school districts with Hispanic enrollments of at least 5 percent (Farley, *Recent Trends in School Segregation*, pp. 32-34). As noted earlier, however, such a measure is probably not germane to the effects of desegregation on test scores.

112. Selection changes have often been subsumed in the broader category of changes in the composition of the test-taking group. Compositional changes stemming from selection changes, however, have very different implications from those reflecting changes in the characteristics of the school-age population. For that reason, changes in the ethnic composition of the group tested that can be considered selection changes are discussed here rather than with changes in the ethnic composition of the entire cohort (discussed above under societal factors).

Retention

Although changes in retention have often been cited as having contributed to the decline in achievement, they had little or no direct role during much of the period of the decline (from about 1968 to the end of the 1970s). Earlier sizable increases in retention, however, could have contributed indirectly to the decline. One hypothesis, for example, holds that schools might have gradually lessened academic demands during the 1970s in response to the earlier increases in retention and that this "pedagogical echo" continued to contribute to the decline of test scores for years after the retention changes themselves ended. 113/ Retention changes also did not contribute appreciably to the rise in test scores and may have impeded it slightly in some instances.

Achievement trends among students under age 16 have been largely unaffected by changes in retention because of mandatory attendance laws. 114/ The issue is the extent to which trends in achievement among older students--primarily high school juniors and seniors--can be attributed to such changes. Thus, the most relevant available measures of changes in retention are the proportion of 16- and 17-year-olds enrolled in school below the college level, and the proportion of youth in each of those age groups enrolled in the modal grade for their age (that is, 16-year-olds enrolled as juniors, and 17-year-olds enrolled as seniors). To the extent that testing is linked to grade rather than age, the latter measure is superior. 115/

---

113. William W. Turnbull, *Student Change, Program Change: Why SAT Scores Kept Falling* (New York: College Entrance Examination Board, 1985).

114. Retention rates have not been constant in earlier grades, but the changes have been much smaller than in the higher grades. For example, between the cohorts that entered fifth grade in 1954 and 1964, the retention rate increased by almost 11 percentage points in the eleventh and twelfth grades but by only 3 percentage points in the eighth grade. National Center for Education Statistics, *Digest of Education Statistics, 1982* (Washington, D.C.: Department of Education, 1981), Table 9.

115. The relevant data reflect students' ages in October, so most students graduating at the age of 18 and all graduating at the age of 17 are included in the category of 16- and 17-year-olds enrolled below the college level.

The more familiar graduation and dropout rates are less germane. For most tests (excluding graduation "exit" exams), the most important consideration is whether the student is present to be tested, not whether he or she graduates.

The Decline in Test Scores. Among all ethnic and racial groups, the proportion of 16- and 17-year-olds enrolled showed some fluctuations but little net change between 1968 and 1979--the end of the achievement decline among students of that age (see Figure A-3). In contrast, retention had increased sharply from 1950 to the late 1960s. 116/ It then increased slightly in the late 1960s. This upturn was followed by a short-lived dip in retention, after which the rate remained at about its 1967 level until 1979. 117/

Modal-grade enrollment trends are largely consistent with the overall enrollment of 16- and 17-year-olds. Enrollment of 17-year-olds as high school seniors rose considerably some time between 1964 and 1969, although the form of the available data--a single average for 1964, 1965, and 1966, and annual data beginning in 1969--make it impossible to pinpoint more precisely when that increase occurred (see Figure A-4). 118/ That rise in enrollment antedated much of the achievement decline. The increase eroded quickly, however, and enrollment then vacillated slightly until the end of the 1970s.

The enrollment of 16-year-olds as high school juniors was slightly more consistent with trends in test scores, since there were hints of an enrollment increase during the latter 1970s--the final years of declining scores in that age group (see Figure A-4). This increase, however, was

---

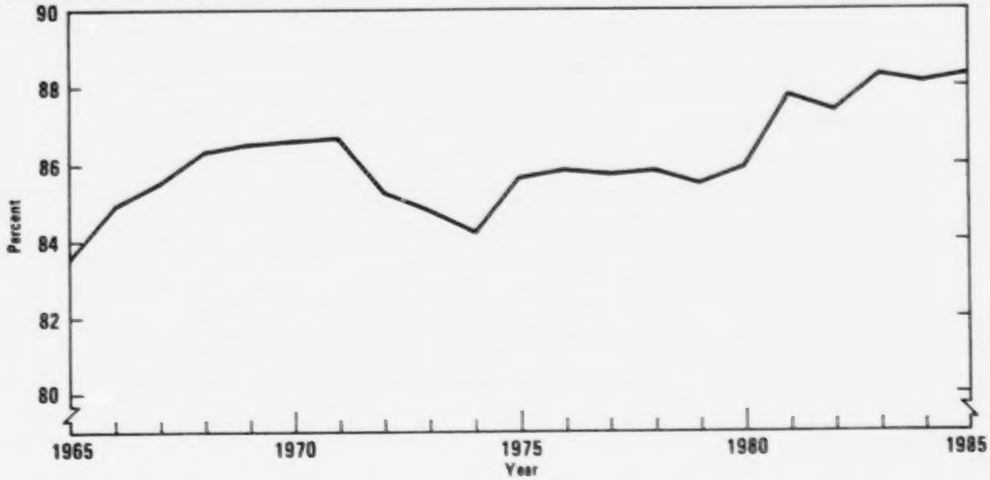
116. The measure of retention available for this earlier period is somewhat different: the proportion of those enrolled in grade five who remain enrolled until grade twelve--or graduate--seven years later. These proportions increased markedly until the fall of 1968 but showed little change after 1968. (National Center for Education Statistics, *Digest of Education Statistics, 1982*, Table 9.)

117. In contrast, the retention rate among black students increased, though erratically and only modestly, between 1967 and 1979--from 83 percent to 87 percent. Given that black students comprised only about 14 percent of the age group in 1979, however, this small change in their retention rate would have contributed only trivially to the test score decline in the age group as a whole.

118. Annual data before 1969 are not currently available (Paul Siegel, Bureau of the Census, personal communication, May 1987).

Figure A-3.

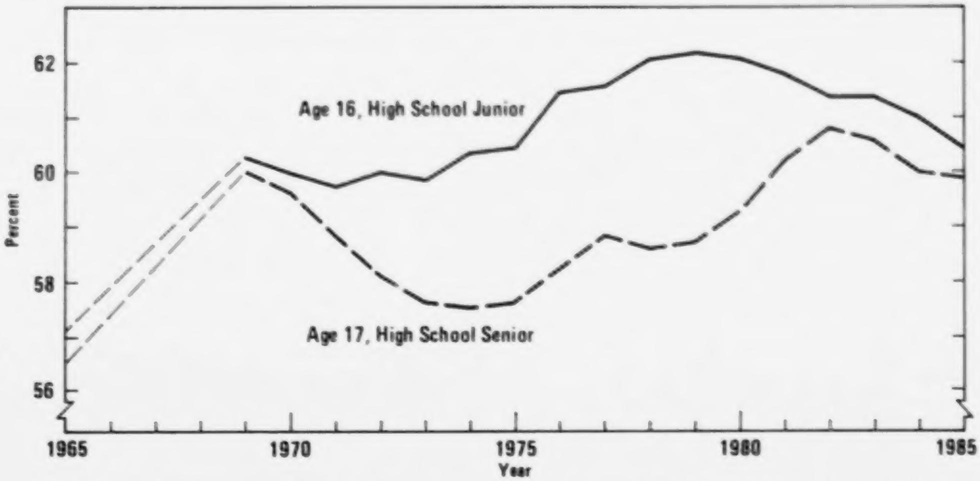
Percent of 16- and 17-Year-Olds Enrolled Below College Level (All ethnic groups combined)



SOURCE: Bureau of the Census, *School Enrollment: Social and Economic Characteristics of Students*, Series P-20 (Washington, D.C.: U.S. Department of Commerce, various years).

Figure A-4.

Percent of Age Group in Modal Grade (Three-year moving averages)



SOURCE: Congressional Budget Office calculations based on Bureau of the Census, *School Enrollment: Social and Economic Characteristics of Students*, Series P-20 (Washington, D.C.: U.S. Department of Commerce, various years) and unpublished data.

NOTE: 1969 value is for single year. No data are available for 1966 through 1968.

trivial and erratic until the last few years of declining scores, and even in its entirety would have had only a slight effect on average test scores. 119/

The Subsequent Rise in Scores. Evidence about retention trends since the late 1970s, when test scores started rising in the senior high grades, is less consistent, but it is nonetheless clear that retention did not contribute appreciably to the rise.

Overall pre-college enrollment of 16- and 17-year-olds has risen modestly--about three percentage points--since the late 1970s (see Figure A-3). Because students at risk of dropping out generally are low achievers, this trend suggests that the recent upturn in test scores among high school students might have been slightly larger if the increase in retention had not occurred.

Recent trends in modal-grade enrollments are somewhat different but also do not indicate a major contribution to the rise in test scores. The proportion of 17-year-olds enrolled as seniors rose after 1978, fell a bit after 1982, but remained higher in 1985 than in 1978 (see Figure A-4). If this slight change had any effect, it would have impeded the rise in scores very slightly. The proportion of 16-year-olds enrolled as juniors has fallen since 1979, but the drop was so small that it could have contributed only slightly to the rise in scores in that grade.

Self-Selection

Changes in the self-selection of students pertain only to college admissions tests such as the SAT and the ACT, but it is an important factor because these tests are among the most salient in the debate about educational achievement. Furthermore, these changes contributed substantially to the decline in both SAT and ACT scores, thereby exaggerating the deterioration of overall achievement. 120/

---

119. For example, between 1972 and 1978, the proportion of 16-year-olds enrolled in the modal grade increased from 60 percent to 62.1 percent. If one assumes that the students newly retained scored on average a full standard deviation below the mean of other students, the effect of this change would be to lower the overall average test score by about 0.02 standard deviation.

120. Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination* (New York: College Entrance Examination Board, 1977); L. A. Munday, *Declining Admissions Test Scores* (Iowa City: The American College Testing Program, 1976).

The effects of self-selection on SAT scores have been particularly well evaluated. The extent of these effects during part of the decline (until 1971) was estimated in one study by comparing the reading comprehension of nationally representative samples of high school seniors and college entrants to those of students taking the SAT. 121/ This study offered two different estimates of the impact of selection. Both estimates indicated that selection--in this instance, the growing number of less able students choosing to take the test--roughly doubled the apparent size of the decline among students taking the SAT during those years. It exaggerated the drop in scores on the SAT-Verbal by about 75 percent and the decline in the reading comprehension scores of students taking the SAT by about 125 percent (see Table A-2). 122/

The continuing decline of SAT scores after 1971, however, probably was not caused by changes in self-selection. Between 1971 and 1976--that is, until just a few years before SAT scores stopped declining--the proportion of high school graduates taking the test fell slightly (see box, Chapter II). This decrease should have made the test-taking group more select and thus would have worked against the continuing decline in scores. The proportion of test-takers classifying themselves as white declined, but that drop paralleled the corresponding growth in the nonminority share of the high school cohort as a whole, indicating that the growth in the

---

121. Albert E. Beaton, Thomas L. Hilton, and William B. Schrader, *Changes in the Verbal Abilities of High School Seniors, College Entrants, and SAT Candidates between 1960 and 1972* (New York: College Entrance Examination Board, June 1977). The estimates given here reflect selection changes affecting students taking the SAT above and beyond those that affected the student population as a whole (such as trends in retention). This discussion thus differs from certain other assessments of SAT trends that consider both types of selection changes together and estimate a larger impact of "compositional" changes on SAT scores during the early years of the decline. (Compare Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination*, Part Three.)

122. The exaggeration of the SAT decline by selection was partially offset by "scale drift"--a gradual lessening of the SAT's difficulty caused by inadvertent errors of equating--that moderated the total decline in scores. The estimates here reflect the extent to which the observed decline--augmented by selection but diminished by scale drift--overstated the true drop attributable to ability changes. The underlying effect of selection (which would be apparent if the errors of equating were corrected) is considerably larger but is not relevant here. The greater impact of selection on reading comprehension scores, in comparison to SAT scores, might in part reflect the lack of scale drift on the reading comprehension test.

TABLE A-2. TRENDS IN SAT-VERBAL SCORES AND READING COMPREHENSION AMONG ALL HIGH SCHOOL SENIORS, AND SAT CANDIDATES, 1959 TO 1971

Test	Group	Decline (In standard deviations)
SAT-Verbal		
Observed total decline a/	Students taking the SAT	.19
Change in ability only b/	Students taking the SAT	.11
Reading Comprehension	All seniors	.16
	Students taking the SAT	.36

SOURCE: Adapted from Albert E. Beaton, Thomas L. Hilton, and William B. Schrader, *Changes in the Verbal Abilities of High School Seniors, College Entrants, and SAT Candidates Between 1960 and 1972* (New York: College Entrance Examination Board, June 1977).

NOTE: The years 1960 and 1972 in the cited source refer to the springs of those years; the labels on this table refer to the 1959-1960 and 1971-1972 school years for consistency with other cited sources.

a. In the study sample only. The national decline was 0.21 standard deviation.

b. Estimate of score change after removing the effects of both selection and scale drift.

nonminority share of students taking the SAT did not represent a change in self-selection. 123/

Changes in self-selection also appear not to account for the recent upturn in SAT scores; indeed, they might have impeded it, perhaps substan-

---

123. Since the Student Descriptive Questionnaire on which these estimates are based includes "Mexican-American" and "Puerto Rican" as explicit choices, the "white" category can be considered non-Hispanic white and thus corresponds closely to the nonminority category used here as a comparison.

tially. Since 1976, the proportion of graduates taking the test has risen sharply (see box, Chapter II). This increase probably made the test-taking group less select and therefore probably hindered the rise in average scores. Changes in the ethnic mix of the test-taking group since scores began rising--which reflect both self-selection and trends in the composition of the cohort as a whole--also probably made no substantial contribution to the rise in average scores. For example, the slight and erratic decline in the share of black students in the test-taking group, which represents a trend in self-selection, most likely accounts for roughly 0.2 points of the rise in average scores between 1979 and 1984--well under half a percent of the total increase.